Author:
Giles, David Clifford

Title:
Visual memory and spelling in 13 year olds.

"Visual memory and spelling in 13 year olds"

David Clifford Giles

A thesis submitted through Cheltenham & Gloucester College of Higher Education to the University of Bristol in accordance with the requirements of the degree of psychology in the Faculty of Social Sciences.

October 1996

Abstract

Although it is widely claimed that visual memory plays an important role in spelling, most of the claims arise from neuropsychology and the analysis of spelling errors. Little empirical evidence can be found in experimental psychology to support the argument. The literature on visual memory and *reading* suggests that the apparently superior visual recall of good readers owes more to their ability to apply verbal labels to test items than to superior visual memory. In this thesis a selection of visual sequential memory, visual recognition, and naming speed tasks are devised and three major studies are undertaken to determine the relationship between performance on these tasks and the spelling of 13 year old schoolchildren. Performance on the tests of visual sequential memory and visual recognition did not produce significant differences between poor spellers and controls, but the study of individual subjects suggests that the tests might highlight cases where spelling difficulties can be attributed to specific visual memory impairment. On only one task - the Animals test - do controls achieve significantly higher scores than poor spellers of average intelligence. This test combines visual sequential memory with verbal labelling, and it is suggested that it may be the combination of skills, rather than either visual memory or verbal labelling *per se*, that handicaps poor spellers. The results are discussed with reference to general theories of lexical representation, working memory, strategy use and the dyslexic automatisation deficit hypothesis of Nicolson and Fawcett (1990).

Acknowledgements

Many thanks to my supervisor, Colin Terrell, for his hard work, inspiration and support over the last three years; to my second supervisor Barry Torrington, for his useful comments; to Clare Wood and Terri Passenger for many profitable discussions; to Dai Jones and Pav Chera for showing me how to switch on a computer, and to Jonathan Elcock for helping me out with hard sums. Thanks also to the members of staff at schools involved in the project who interrupted their busy schedule to accommodate my research, and of course to the many children who participated in the study.

Author's declaration

The work contained in this thesis is all the candidate's own work. The views expressed in the dissertation are those of the author and not of the University.

Signed ....David Giles.................

TABLE OF CONTENTS

CHAPTER ONE: THE ROLE OF VISUAL MEMORY IN SPELLING

CHAPTER TWO: VISUAL MEMORY TESTING

CHAPTER THREE: VISUAL MEMORY, VERBAL LABELLING AND READING

CHAPTER FOUR: VISUAL SEQUENTIAL MEMORY: TEST CONSTRUCTION AND PILOT STUDIES

CHAPTER FIVE: VISUAL SEQUENTIAL MEMORY AND SPELLING ABILITY

CHAPTER SIX: VISUAL RECOGNITION AND NAMING SPEED: TEST CONSTRUCTION AND PILOT STUDIES

CHAPTER SEVEN: VISUAL RECOGNITION, VERBAL LABELLING AND SPELLING: A CORRELATIONAL STUDY.

CHAPTER EIGHT: VISUAL MEMORY, VERBAL LABELLING AND SPELLING: A SPELLING-LEVEL DESIGN STUDY.

CHAPTER NINE: GENERAL DISCUSSION

LIST OF TABLES AND ILLUSTRATIVE MATERIAL

CHAPTER ONE: THE ROLE OF VISUAL MEMORY IN SPELLING

1.1 Introduction to the literature review chapters

The first three chapters of the thesis set out to review previous research on the subject by discussing relevant literature from three perspectives.

i) The first chapter begins by discussing a number of models of the cognitive processes involved in spelling which have been developed, mostly on the basis of neuropsychological case studies. There follows a review of the experimental literature linking spelling to visual memory. Few empirical studies have been undertaken, although several studies have made such links from analysis of spelling errors, and from the spelling performance of hearing-impaired and visually-impaired groups.

ii) The second chapter concentrates on visual memory itself, with particular regard to the methodologies involved in assessment. The experimental evidence for different aspects of visual memory is discussed, and there is an analysis of the types of stimuli which have been used in the testing process. Specifically, the issue of verbal labelling is discussed, with particular reference to the strategies researchers have adopted to try and control it.

iii) The third chapter looks at the way in which visual memory has been researched with regard to its involvement in children's reading. Several studies are discussed which relate reading deficiency and "dyslexia" to poor visual memory. There follows a review of studies in which these group differences can be explained by verbal labelling of the stimuli rather than visual memory *per se*. Finally there is a review of studies which examine verbal labelling as a factor by itself - picture naming and rapid automatised naming - and its relation to reading ability.

1.1.1 Introduction to Chapter 1

If asked to copy from dictation the sentence "the sheep was in the field", how does one know how to represent the sound /ʃi:p/? The spelling *sheap* would be acceptable according to the sound-spelling rules of English, as would *shepe*. Yet competent spellers have no apparent difficulty assembling the correct letter string. Nor do they have too much trouble spelling words like *yacht* and *phlegm* which deviate markedly from English orthographic conventions.

There seems little doubt that, at some stage of the spelling process, visual memory must play a part. Given that visual memory is likely to vary among individuals, a topic for investigation is whether someone with a poor visual memory is necessarily a poor speller. It may be that there are too many other factors involved for this relationship to hold true. This thesis sets out to explore these questions empirically.

In this opening chapter, it is argued that the existing literature concerning visual factors in spelling is far from conclusive. The assertion that visual memory contributes to spelling is stated by numerous authors (e.g., Lennox and Siegel, 1994; Cornelissen, Bradley, Fowler and Stein, 1994; Goswami, 1992; Huston, 1991; Ellis, 1984) but is, as yet, not demonstrated empirically. Little experimental work has been undertaken which compares visual memory *per se* among individuals in relation to spelling ability. Until this gap in the literature has been closed it is difficult to ascertain the precise role played by visual memory in spelling.

It is proposed to investigate the role of visual memory in spelling from four basic perspectives. Firstly, from the perspective of cognitive and developmental modelling. Various attempts to model the spelling process are discussed, with particular reference to the components of these models that might represent visual memory. As yet, no model of spelling has been constructed in which visual memory is deemed to play an explicit role. However, each model discussed contains a component in which "orthographic representations" are said to be stored, and there are a number of neuropsychological case studies which have been cited as evidence to support the existence of such a component. It is suggested that the representations stored in this component may involve the visual system in both storage and recall..

Secondly, there is a review of the experimental literature which has examined the relationship between visual memory and spelling. This literature is not extensive, and its findings are not conclusive. However the small number of experimental studies can be split broadly into two groups, one of which examines the role of visual recall of sequential information, and the other which examines visual recognition memory. It is suggested that these are two distinct processes, either of which may be of importance with regard to spelling.

The third perspective is the imputation of visual memory processes from the analysis of spelling errors. Two major studies are reviewed, each of which examines a large corpus of spelling errors of different types of word. It is suggested that visual memory may be more important for the spelling of phonologically irregular words than for the spelling of regular words. There is also a discussion of the literature relating to subgroups of dyslexia, in which it is argued that there are at least two discrete types of spelling disability, one of which appears to involve a deficit at the level of visual memory.

The fourth perspective is the spelling of hearing- and visually-impaired individuals. If visual memory plays an important role in the spelling process, the spelling of the born deaf should display the same level of competence as the spelling of hearing individuals, although with no sensitivity to phonological information. Likewise, the spelling of individuals with visual impairments should display a considerable deficit and an over-reliance on phonological information. While the literature on deaf spelling is inconclusive on this issue, one major study on the spelling of the visually-impaired is reviewed which appears to support the argument for the importance of visual memory in spelling.

1.2.1 Cognitive models of spelling

In this section, the dual-route model of spelling is discussed, in which two possible ways of spelling a word are proposed. One is a sound-to-spelling route, in which spellings are pieced together on the basis of phonological information and the application of sound-spelling rules; the other is a direct route in which spellings are retrieved from a putative "internal lexicon". A number of case studies are reviewed which appear to lend weight to the existence of an internal lexicon. The question of relevance to this thesis is *how* these spellings are represented in this hypothetical store. It is suggested that these representations may be visual in nature. However, the visual nature of representation has not been examined in any depth by these authors. Computational models of spelling are also discussed briefly; these models also fail to specify how spellings might be represented internally.

1.2.1.1 The dual-route model of spelling

Morton's (1980) model of the reading and spelling process has emerged as the prototype for a series of similar models (e.g., Link and Caramazza, 1994; Ellis and Young, 1988;

Shallice, 1988; Ellis, 1984), collectively known as the dual-route model of spelling (Barry, 1994). These models take as their initial premise the idea that a substantial number of English words cannot be reliably assembled using phonological information. "*It is particularly important for a language like English that spellings should be retrieved as wholes from memory, because of the presence in the language of so many words with unpredictable, irregular spellings.*" (Ellis and Young, 1988, p.227)

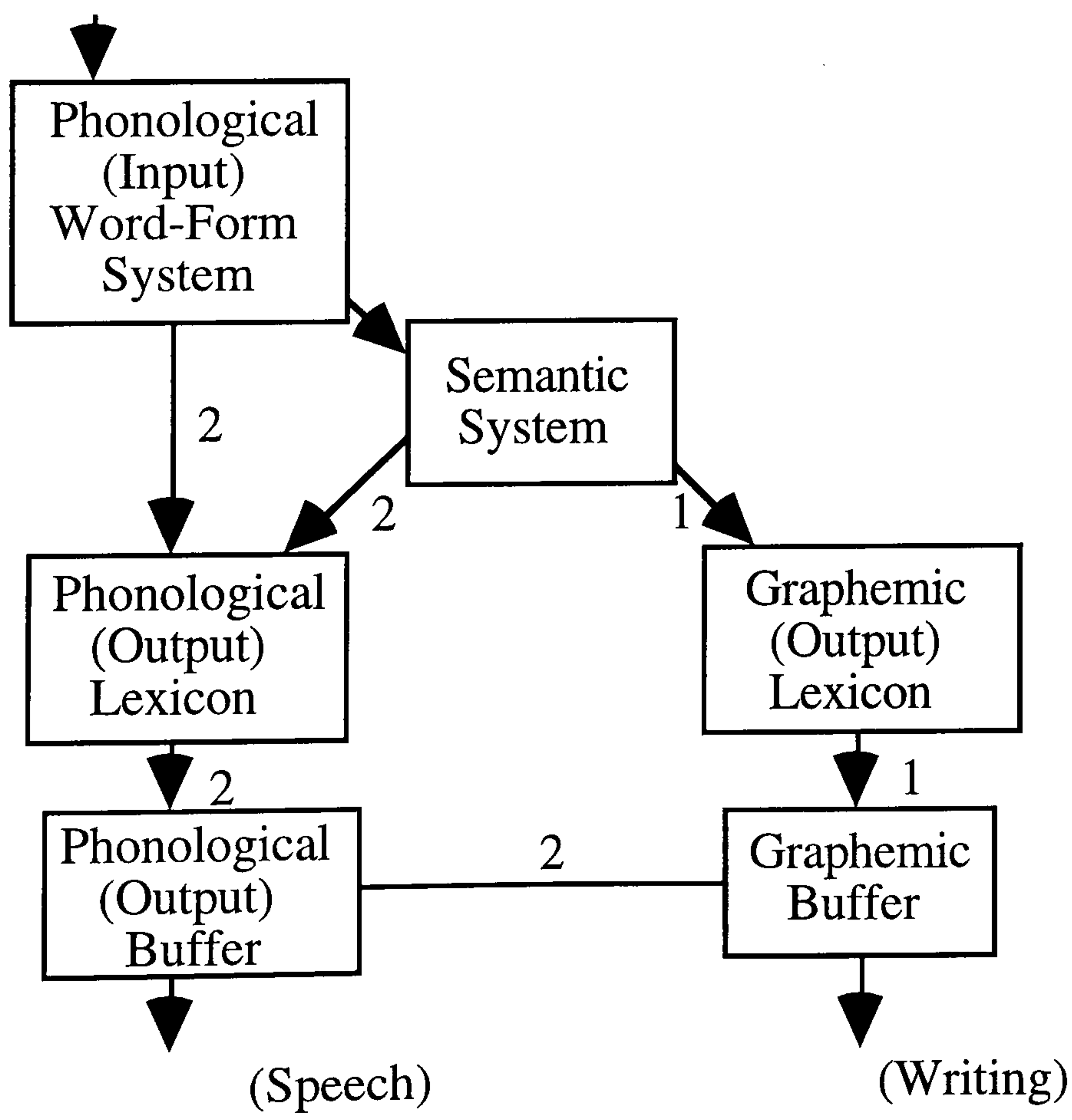
Cohen and Squire (1980) suggested that memory operates on the basis of declarative (knowing what) knowledge and procedural (knowing how) knowledge. From this theory it could be said that the phonological route uses procedural, rule-based knowledge, while the lexical route uses stored factual (declarative) knowledge about the precise spellings of words.

It has been estimated (Rudorf, 1965) that only 50% of high-frequency English words can be spelled accurately using a strictly phonological approach. Given the number of plausible phonological alternatives for many spellings (as in the example of *sheep* in section 1.1), it appears that well over half of all English spellings require some degree of visual recall to supplement the phonological information. The same has also been shown to be true of the French language (Véronis, 1988).

Therefore, the dual-route model allows for spellings to be pieced together using a sound-spelling (phonological) route, and also for spellings to be retrieved from a putative orthographic store (the lexical route). It is argued that both routes are used by normal spellers in the spelling of most words (Barry, 1994), although the model has been developed to account for the spellings of brain-damaged individuals. Case studies have been described which suggest that damage to one route may lead a patient to rely wholly on the other route for spelling (Ellis and Young, 1988; Shallice, 1988).

The manner in which orthographic information is represented in the lexical spelling route is the subject of a great degree of speculation. Morton (1980) and Shallice (1988) use the term "graphemic output lexicon" to describe the store of graphemic representations of known words which enables a speller to construct words without requiring phonological information. Link and Caramazza (1994) postulate a similar structure, which they term the "orthographic output lexicon". In all dual-route models, this structure feeds into a "graphemic buffer", in which these activated graphemic representations are held during the act of writing. The concept of a graphemic buffer dates back to Schaffer's (1975) model of typing skill, in which this structure was said to preserve the ordinal letter information. Figure 1 displays the section of Shallice's (1988) simplified version of the dual-route model.

Figure 1. A simple dual-route model of spelling (from Shallice, 1988). 1 is the lexical/semantic route; 2 is the phonological route using sound-to-spelling rules



It has been suggested (Barry, 1994) that the lexical and phonological routes may operate in tandem for most words. The theory suggests that consonants are represented phonologically, but that vowels are represented lexically. In other words, we use our phonological knowledge to recall the 'y' and 't' of *yacht* but the phoneme in the centre ('o') needs to be represented lexically (visually, as 'ach'). However, this theory fails to work for a word such as *ceiling*, with an ambiguous first phoneme.

1.2.1.2 Evidence from neuropsychological case studies

In the last fifteen years the theoretical foundations of the dual-route model have been strengthened by the description of numerous single case studies of brain-damaged individuals, each of whose impairments appeared to affect one or other spelling route.

Shallice (1981) describes a patient PR, who seemed to have a selective inability to spell nonwords. Accuracy of nonwords was only 18%, compared with 91% for known words. He had no apparent deficit in auditory-verbal memory. Thus it appeared that his specific deficit concerned his ability to piece together unknown words from the given phonological information using sound-spelling rules. An example of a patient whose impairment was at the level of the "graphemic output lexicon" was PT of Hatfield and Patterson (1983). On words with regular spellings, PT displayed 93% accuracy, compared with 38% on irregular spellings. *Spade* was spelled 'spaid' and *flood* 'flud'. This individual appeared to have a deficit which prevented him from activating stored representations of known words.

Bub and Kertesz (1982) also report a patient, MH, who was unable to spell nonwords; her knowledge of sound-spelling rules was impaired, with the result that she appeared to rely on the graphemic output lexicon to spell. Therefore she made many clearly "visual" errors such as "orchatria" for orchestra. Another single case study, of an aphasic reported to have no inner speech (Levine, Calvanio, and Popovics, 1982), appears to lend further support for the existence of the lexical route of spelling. This patient, though able to spell many words correctly, nonetheless produced some bizarre errors such as "string yard bed" for *hammock* and "find way in/out puzzle" for *maze*. In a rhyming matching task he could match words only if they were spelled alike (e.g. *bee/tree*), failing to match *soap/rope* or *bear/chair*. Goodman and Caramazza (1986) also describe a patient whose bizarre spelling errors indicate "*a selective deficit at the level of the Graphemic Output Lexicon*" (ibid., p.288-289).

If an internal store of familiar words exists, unaffected by phonological information, then we would expect mis-spellings to be consistent across time. Both Campbell (1987) and Goulandris and Snowling (1991) describe case studies of undergraduate subjects whose consistent mis-spellings are said to provide evidence of visual memory deficits. However, the examples provided in these studies, e.g. "logicle" for *logical* and "garantee" for *guarantee* would appear to be the most likely spellings to be assembled using sound-spelling rules. While the principle of error consistency would seem to be essential to support this theory, examples of consistent *nonphonetic* errors are required too.

While the findings from these single-case studies provide evidence of the role of the visual system in the spelling of specific individuals, there are difficulties in generalising these findings across the whole population. Analyses of deficits following brain damage may highlight specific systems that are used in cognitive processes, but they shed little light on the way those systems operate in normal processing.

1.2.1.3 The nature of orthographic representation

For the purpose of this thesis, the most important question raised by the dual-route model of spelling is *how* the internal lexicon is said to represent spelling information. This question has not been addressed directly by any of the authors.

Link and Caramazza (1994) suggest that graphemic representations constitute "*multi-dimensional objects specifying independent types of orthographic information*" (*ibid.*, p. 268-269). They argue that each letter is represented according to its spatial position within the word. These representations are activated, then stored in the "graphemic buffer" during the act of writing. Although the authors do not make the inference explicitly, it could be argued that these "objects" are represented visually, as stored memories of words. However, their model fails to account for the way in which pictograms are represented in a language such as Mandarin. Indeed, as they argue, even the dual-route model cannot explain spelling in all types of orthography.

Shallice's (1981) patient PR, who was described in section 1.2.1.2, seemed to have a specific deficit located in the phonological spelling route. Therefore, he relied on the graphemic output lexicon to spell words. His experience of spelling was described as "*transcribing from an inner screen on which he sees the word*" (*ibid.*, p.418). That study described the experience of one individual who had an over-reliance on visual memory. It is impossible to generalise on the basis of a single case, although a similar process might form *part* of the general spelling process, especially for phonologically irregular words. Shallice (1988, p. 156-157) says: "*Possibly, the contents of the graphemic buffer can be fed by a process related to imagery into a visual short-term store, held there, and* *transferred back, at least by some subjects. We know far too little about the writing process in normal subjects to be able to assess whether this is possible.*" This thesis sets out in the hope of providing some information that might advance knowledge in this area.

1.2.1.4 Computational models of spelling

In this section, there will be a short discussion concerning recent attempts to model the cognitive processes involved in spelling through computer programming (Brown and Loosemore, 1994; Houghton, Glasspool and Shallice, 1994; Olson and Caramazza, 1994). Although the dual-route model of spelling seems to account for the spelling deficits reported by neurological patients, it is debatable to what extent one can generalise from this type of model to the general population. Computational modelling of cognitive processes has become popular in recent years because it appears to offer a formal expression of the processes under consideration, and a certain degree of explicitness about these processes is required in order to formulate such a model (Brown and Loosemore, 1994). At some point, therefore, the question of visual memory may need to be addressed to account for the manner in which orthographic information is represented.

The three models under consideration all follow the connectionist approach to computational modelling of psychological processes. Briefly, this approach entails simulating the electrical activity of the brain by constructing interconnected networks of artificial neurons (see Rumelhart and McClelland, 1986). These "activate" one another in the same manner as real neurons, each being either excited or inhibited by the information received from its neighbours. This is achieved by ascribing weights or strengths to each artificial neuron, or sub-population of neurons. These "neural networks" differ from traditional computational models because they process information in a parallel fashion rather than in a serial fashion (i.e. a series of if/then commands). This means that processing is more flexible, and allows such a model to learn, generalise and respond to novel input (Brown and Loosemore, 1994).

One early attempt to model the processing of skilled typists was undertaken by Rumelhart and Norman (1982), who used a "competitive queuing" approach in which all letters in a word become active in parallel, and the correct serial order is selected on the basis of the typist's hand position. Houghton, Glasspool and Shallice (1994) have applied the same approach to a connectionist model of normal spelling. In this case it is not letters that become activated, but phonemes. The authors themselves admit that this model is "*only a starting point*" (*ibid.*, p.400), but already the limitations of such a model have become clear, in that it only offers an adequate explanation for "typos", or slips of the pen of normal spellers. It does not account for phonologically plausible errors, such as the "Good Fonetic (*sic*) Equivalents" described by Boder and Jarrico (1982), because it has no phonological input. Nor does it make any attempt to account for how orthographic representations (i.e. known words stored in an internal lexicon) become available. "*We are not concerned here with where activation of the GOL [Graphemic Output Lexicon] comes from*" (*ibid.*, p. 383). Failure to address the roles played by either phonology *or* visual memory would therefore appear to be a critical flaw in this model.

The "NETspell" model of Olson and Caramazza (1994) is more successful, in that it has successfully simulated the phonologically plausible errors of neurological patients with deficits affecting the lexical spelling route. However it has not attempted to model the lexical spelling route itself. Therefore it cannot spell homophones or phonologically irregular words.

Brown and Loosemore (1994) have attempted to model the processes involved in learning spelling, in an attempt to simulate developmental dyslexia. They found that, by damaging the model to differing degrees it was possible to retard the learning process and produce errors that simulated those of actual dyslexic spellers. However, the emphasis of this model was on slow learning of sound-spelling rules and not on the acquisition of an internal lexicon. Therefore it is constrained by the particular theory it uses (Frith, 1985, discussed in the next section), and again makes no attempt to account for visual memory.

A general criticism of the models described in this section is that they fail to make any attempt to simulate the role that visual memory might play in the spelling process. One general criticism of all connectionist models is that they tend to be "technology driven" (Houghton, Glasspool and Shallice, 1994, p. 371), and it may well be easier to simulate the learning of sound-spelling rules than the considerably more complex nature of visual representation. In order for a computational model to fully simulate the spelling process it may need to incorporate a simulation of the visual pathway, in order to account for all the possible influences on the normal human brain's ability to retrieve the correct spelling of a word. Such a model would be extremely ambitious; indeed, even Brown and Loosemore (1994) admit that the complex nature of spelling may well be beyond the limits of computer programming.

1.2.2 Developmental models of spelling

In this section, a different approach to modelling spelling is considered, albeit one which played an influential role in the Brown and Loosemore (1994) model described in the previous section. This is the stage model of reading and spelling development of Frith (1985), which attempts to account for the pattern of spelling errors made by normal children as a consequence of maturational factors. It also attempts to account for developmental dyslexia, by specifying stages at which the spelling process might be retarded.

This model proposes three successive stages for reading development: the *logographic* stage, where words are read on a look-and-say basis; the *alphabetic* stage, where words are decoded using sound-spelling rules, and the *orthographic* stage, where readers begin to use direct access to stored representations of known words, and analogies with familiar words to decode unknown words. By analogy to the dual-route model of spelling discussed in section 1.2.1, this stage could be regarded as the point at which the "graphemic output lexicon" begins to exert a major influence on spelling. For the purpose of this thesis, it could be regarded as the stage at which visual memory becomes important. Frith suggests that in normal circumstances, the orthographic stage begins at around the age of 12.

According to the stage model, children learn to spell using sound-spelling rules, by-passing the logographic stage. This argument is based on the work of Bryant and Bradley (1985, 1980), who found that 6-7 year old children were capable of spelling words like *leg* and *cut* despite being unable to read them. However, the orthographic stage in spelling is not reached until around three years after it is reached in reading. Frith describes this stage as "*the instant analysis of words into orthographic units without phonological conversion...they are internally represented as abstract letter-by-letter strings*" (ibid., p. 306). This theory is based partly on the existence of a group of children who are good readers but poor spellers (Frith, 1980). These children were said to be reading using the compensatory process of "partial cues", rather than the full orthographic representations required for spelling. Alegria and Mousty (1994) suggest that the use of partial cues in reading "*results from the subject's inability to store full cues as a consequence of a general visual memory deficit*" (*ibid.*, p. 223).

Other authors also argue for the importance of visual memory in the later stages of learning to spell: "*Initially the scales are tipped in favour of phonological skills when children begin* *to spell, but this situation changes with time as children increasingly make use of visual memories for letter strings which they glean through reading*" (Cornelissen *et al.*, 1994, p.717). "*It is phonological rather than visual skills that play the greatest role in spelling* development, *even though visual memory for spelling patterns will be important for spelling proficiency*" (Goswami, 1992, p.967).

Unfortunately Frith herself is not explicit about the ways in which these internal letter-by-letter strings might be represented, how they might be accessed, or whether different types of word are stored differently. Curiously, she describes the orthographic stage as "non-visual" (Frith, 1985, p. 306).

1.3 Visual memory processes involved in spelling

In this section, some experimental studies are reviewed which examine specific visual memory processes which might be involved in spelling. In the models discussed in the previous sections, the visual processes involved in spelling have only been *inferred* from the general, rather abstract, notion of internal "orthographic representation". Most modellers agree on the existence of an internal lexicon from which known spellings are retrieved, and it has been argued that this lexicon becomes increasingly important with age (Frith, 1985; Goswami, 1992). But the precise processes by which known spellings are retrieved have not been examined in any depth.

This section attempts to collate the small number of experimental studies which argue for the role of visual memory in spelling, by proposing three separate visual memory processes: visual sequential memory, visual recognition, and visual imagery. These processes are considered also with reference to the various models of spelling discussed previously.

1.3.1 Visual sequential memory

When spelling a word such as *yacht*, it is possible that skilled spellers process the letter string as a series of discrete symbolic units. "*The element of phoneme to grapheme contrast also requires sequential scanning and sequential memory*" (Thomson, 1984, p.144-45). Shallice (1988) argues that this process takes place while the stored representation of a word is held in a "graphemic buffer". The "competitive queuing" model of Houghton, Glasspool and Shallice (1994) attempts to simulate this process.

Two studies appear to have demonstrated the process experimentally. Jensen (1962) examined over 1500 spelling errors made on a selected sample of words. He found that the serial position curve for the errors in these words corresponded closely to the serial position curve for errors in a separate rote learning task. The rote learning task required subjects to recall a series of coloured geometric shapes. In the first condition these shapes were presented individually for three seconds each; in the second condition they were presented as a simultaneous sequence. It was found that the error curve produced in the second condition was more closely related to the spelling error curve, leading the author to conclude that the second task was analogous to the memory processes involved in spelling.

On the surface, this appears to be a plausible interpretation of the data. However, there are two methodological flaws which undermine the argument. Firstly, although the author recognises the potential variability of spelling errors for different types of words, this factor is not considered in the experimental design, and the serial position curve may be hiding wide variations. There is a substantial difference between, say, omitting a silent letter and substituting an incorrect grapheme .

The second flaw concerns the subjects themselves. Although the spelling data was obtained from carefully selected groups of schoolchildren, the rote learning information was obtained from college students. Consequently there is no actual relationship between *performance* on the two types of task, and so any statistical relationship must be at best speculative.

However, subsequent studies on the serial position of spelling errors appear to support the finding. Wing and Baddeley (1980) examined a large corpus of the spelling errors of normal subjects, and found that most errors occurred in the medial positions of words rather than at either end. Similar data was found by Caramazza, Miceli, Villa and Romani (1987) in brain-damaged patients who appeared to have a deficit at the level of the "graphemic buffer" in the dual-route model. It seems to suggest that the primacy and recency effects normally associated with serial recall are also true with respect to spellings.

A more recent study of visual sequential memory in spelling is Bryant and Bradley (1981). Here, 62 10 year old "backward spellers" were matched with both chronological age and spelling age controls on a number of tasks analogous to reading and spelling. The "visual memory" condition required each child to reproduce a nonword, using letter cards, after a 5 second delay. High correlations (above 0.7) were found with spelling age across all the groups. A possible criticism of this study is that the memory task is perhaps too closely related to the spelling process itself to tell us anything about the subjects' visual memory *per se*.

It is surprising, therefore, that no studies have been conducted comparing scores on a spelling test with recall for a visually presented series of abstract shapes. If the graphemic buffer works in the way that Shallice (1988) suggests, then one might argue that a poor speller should display the same difficulty with processing any list of visually-presented items as with processing a string of letters.

1.3.2 Visual recognition

The process of visual sequential memory is essentially one of recall of information, and it is easy to see how it might be important for spelling. The role of visual recognition in spelling is less obvious, and brings into question the nature of internal representation of words.

One way in which visual recognition is used in spelling is as a check. Once a familiar word is written down it can be matched with the internally stored spelling of that word. Therefore, if a person has poor general visual recognition, it could be argued that such a person would find it hard to identify mistakes in their spelling.

A study which examined the visual recognition skills of good and poor spellers was that of Ormrod (1985). Here, ten pairs of good and poor spellers were compared on a short-term recognition task using 9-letter nonwords. The nonwords were generated randomly by a computer program using alternatively vowels and consonants so the nonwords could be pronounceable. They were displayed on a computer screen for one second at a time, and after a one second delay, subjects were required to indicate whether a second word was the same, or different, to its predecessor. When the second word was presented in a different case, or letter size, to its predecessor, both good and poor spellers made significantly more recognition errors. Poor spellers' performance was particularly disrupted by mismatched presentation, falling almost to chance level.

These findings replicated those of Kirsner (1973), who interpreted the importance of visual recognition as indicating that purely phonological representations of written words were insufficient for accurate spelling. Ormrod interpreted her findings in terms of the Frith's "partial cues" hypothesis (Frith, 1980), which argues that poor spellers overlook some of the letters when reading a word, while good spellers read all the letters that they see. Ellis (1994) suggests that good spellers can identify their mis-spellings in order to correct them, but poor spellers lack a "*complete visual orthographic description*" (*ibid.*, p.157) which would enable them to do this.

Some support for this idea comes from Tenney (1980) who, like Ormrod, found that the visual appearance of a word has a significant effect on subjects' ability to choose the correct spelling from two alternatives. In this study, words in the experimental condition were distorted by having each alternate letter appear on one of two vertical lines, in a "zigzag" fashion. Undergraduate subjects made significantly more errors in selecting the correct spelling in the zigzag condition. However, there was no attempt to relate performance on this task to overall spelling ability.

While Ormrod's (1985) study provides convincing evidence for short-term visual memory deficits in poor spellers, the application of all three studies to spelling in general is uncertain. The precise visual appearance of words is unlikely to be related to our knowledge of their spelling patterns; the word *yacht*, for instance, will be seen by a reader in a variety of different contexts - textbooks, fiction, advertisements, television programmes, even shop signs - and in a variety of styles - different cases, colours, sizes and on different surfaces. It would be highly implausible to suggest that there is a representation in the visual memory system for each sighting of a specific word.

An alternative way in which visual recognition might be important for spelling is as an internal check. This process relies heavily on the concept of visual *imagery*, which will be discussed in the next section.

One of the inputs into the graphemic output lexicon in Ellis and Young's (1988) model is from the speech output lexicon, which is intended to account for the use of inner speech while in the act of writing. A study by Sterling (1983) provides some evidence for this process. He undertook a substantial analysis of many different types of spelling error produced in a piece of spontaneous writing by 12 year old children. He attributes many of the children's errors to the observation that they were subvocalising while writing. Hence a word such as *probably* was mis-spelled "probally", since this might be how the writer articulated the word. In the same way, unstressed syllables were often omitted, resulting in errors such as "diffrent" for *different* and "choclate" for *chocolate*. The author argued that in competent spellers, a "lexical monitor" (p. 364) might check the articulated spelling against an internally stored representation and reject it, but where this monitor was defective, the mis-spelling would be allowed through.

Such an explanation is plausible, but treats visual memory as a passive store rather than an active process. It could be argued that a "stored representation" of a spelling needs to be activated in some way for visual recognition to take place. The use of imagery to activate stored representations is the subject of the next section.

One feature of the research discussed in this and the previous section is that, where experimental studies of visual memory have been related to spelling, the stimuli have been predominantly verbal in nature. Given that the majority of poor spellers are also poor readers, it can be argued that they are likely to be at a disadvantage from the outset in studies where verbal stimuli are used. Furthermore, weaknesses in visual memory for verbal information may be distinct from weaknesses in visual memory *per se*. Without a measure of general visual memory obtained using non-verbal materials, it is difficult to glean much from this research about the use of the visual system in spelling.

1.3.3 Visual imagery

As described in section 1.2.1.3, Shallice's (1981) patient PR claimed to spell words as a result of "reading" them off an "inner screen". In the previous sections, the concept of a "graphemic buffer" has been discussed, in which words are "held" while they are being written (Link and Caramazza, 1994). It could be argued that the *generation* of an internal orthographic representation requires the use of some form of visual imagery which enables the speller to "see" the word. This may well be necessary before a word can be "recognised" as such.

There is a considerable literature concerning the topic of visual imagery (see Humphreys and Bruce, 1989). Studies of mental rotation (see Shepard, 1978) and image scanning (e.g., Kosslyn, Ball and Reiser, 1978) suggest that normal subjects are very adept at using an "inner screen" on which to perform cognitive tasks. There has been considerable debate over whether imagery is codified (e.g., Paivio, 1971) or propositional (Pylyshyn, 1973), although, like the dual-route model of spelling, it could be argued that both forms of processing are necessary. There is some neurological evidence for corresponding electrical activity in the visual cortex which suggests that visual imagery is a valid concept (see Farah, 1988). One criticism of visual imagery studies, however, is that they are plagued by experimenter effects (Baddeley, 1986); however, if an experimenter can generate such an effect in a laboratory setting it is surely evidence of a cognitive *ability*, whether it is an automatic process or not.

Nevertheless, the field of visual imagery - indeed of mental representation in general - is riddled with speculation and disagreement. For the purpose of this review, it is intended to select a number of studies whose findings seem to point towards some degree of the use if visual imagery in the spelling process.

One argument for the use of visual imagery in spelling comes from the spelling of nonwords (Campbell, 1985, 1983). In these studies both children and adults, when asked to spell nonwords, were found to use analogies with familiar words with which they had been primed. In the first experimental condition, the experimenter read out a list containing both words and nonwords. Subjects were asked to spell only the nonwords. Each target nonword was preceded in the list by a homophone that was a real word (e.g. *neat*.../fri:t/). In the second experimental condition, the same nonwords were preceded by different homophones (e.g. *feet*.../fri:t/). Controls received only the nonwords.

Both children over the age of 11 and adults showed a significant level of biasing by the respective primed words. This finding was taken to show that, as reading skills increase, children make use of rhyming analogies in order to spell new and unfamiliar words. Campbell also suggests that these analogies are generated by "activation mechanisms" (1985, p.144), whereby the relevant segment is preserved in working memory. Where several items intervened between the prime and the target, vowel spelling consistency was substantially reduced.

This theory of spelling by analogy has been criticised as "computationally opaque" (Goodman and Caramazza, 1986), in that it does not distinguish between lexical and nonlexical (roughly speaking, visual and phonological) processes. How does the effect of analogy work in normal spelling? Is it merely an example of the availability heuristic (Tversky and Kahneman, 1974)? Spelling ability was not a variable addressed in Campbell's studies, so it is not clear how use of analogies might vary among individuals.

A study which produced similar findings to Campbell was that of Seidenberg and Tanenhaus (1979), where subjects were asked to make rhyme judgements on auditorily-presented pairs of words. It was found that it took significantly longer to make rhyme judgements when words were not spelled alike (e.g., *hot-yacht*) than when they were spelled alike (e.g., *hot-cot*). The reverse pattern was obtained for non-rhymes which were spelled alike (e.g., *dead-bead*). These results seem to suggest that subjects were generating internal images of the words during the task.

A more explicit examination of the way in which visual images of words are created is found in the work of Ehri (1980, 1991). She asked 8 year old children to generate internal images of words and then asked them questions about the letters these words contained. She found that the children were as likely to correctly identify silent letters (such as *t* in *listen*) as letters that corresponded to the word's sound. Her explanation is that, "*as the visual forms of words are seen repeatedly, their shape and length are stored and these characteristics create visual spaces in memory for letters to fill*" (1980, p.333-34).

A less reliable measure of visual imagery in spelling was obtained by Walker (1974). Subjects were evaluated on the basis of the Betts Questionnaire Upon Visual Imagery (Betts, 1909), where one is asked to rate on a 7-point scale the vividness of one's mental image generated by a verbal description (e.g. "the sun sinking below the horizon"). A significant interaction was found between visualising ability and the type of errors made in a spelling test (either phonological or visual). High visualisers were more likely to make mistakes on the basis of phonology.

While the nature of internal representation is crucial to the discussion of visual memory and spelling, the subject of visual imagery has proved difficult to study. Therefore it is probably wise to suggest that, while visual imagery must play an integral part in activating stored orthographic representations, means of examining the nature of such imagery are largely unreliable.

1.4.1 Variations within a language

In section 1.2.1.1 it was stated that over 50% of English words cannot be spelled by phonological information alone. Thus it could be argued that the use of the visual system is only necessary when spellings are irregular or ambiguous. Waters, Bruck and Malus-Abramowitz (1988) examined this topic by comparing errors made across a number of defined categories of words. They hypothesised that all subjects would find irregular words harder to spell. Words were classified accordingly:

- *Regular* words, e.g. *must*, which have only one possible graphemic representation
- *Regular** words, e.g. *street*, where the phoneme /i:/ can be represented by *ee* or *ea*
- *Orthographic* words, e.g. *patch*, where knowledge of orthographic conventions is required to insert the letter *t*
- *Morphological* words, e.g. *sign*, whose spelling is dictated by their relationship to another word with the same root (*sign*ature)
- *Strange* words, e.g. *yacht.*

As predicted, the strange words proved hardest to spell. Poor spellers, even by age 12, were only able to achieve a mean score of 7/20 in this category, compared with 16.5/20 for the regular words. Good spellers in this age group had means of 14/20 and 20/20 respectively. The interaction between word category and total number of errors was found to be significant across all age groups (9 to 12). Among poor spellers there was also a significant difference between *regular* and *regular** words.

The findings of this study are not sufficient by themselves to demonstrate the role of visual memory in spelling. The visual memory hypothesis suggests that it ought to be as hard to choose between *bear* and *bare* as it is to choose the correct spelling of *yacht* Ambiguity and irregularity should not necessarily show differing levels of difficulty - it ought to be sufficient to split the words into two categories (regular and irregular) to demonstrate the effect found with five.

In section 1.3.2, it was argued that few studies of spelling compared good and poor spellers on independent measures of visual memory. The few studies that have attempted this type of design include those of Day and Weddell (1972) and Goyen and Martin (1977), who both used short-term visual memory as a variable in studies of spelling error type. Day and Weddell (1972) found a number of error types which correlated with separate measures of auditory and visual sequential memory. Subjects were placed in two groups according to performance on these tests. The "VMS" group comprised subjects who had performed well on the visual memory test but poorly on the auditory memory test, and the "RSR" group were those who had performed poorly on visual memory and well on auditory memory.

The VMS group made considerably fewer single letter omissions than either the RSR group or a control group. They also made fewer "doubling" errors than the other groups (insertion of a single letter for a double, e.g. "coffe" for *coffee*). On another 11 categories of error, however, no difference could be found between VMS and controls, suggesting that these findings may not have theoretical significance as such. There appears to be no consistent explanation as to why either error type should favour those with good visual memory.

In the Goyen and Martin (1977) study, a measure of sequential memory (which involved copying shapes following a 10-second screen projection) failed to load on any factors relating to spelling, although there are insufficient details of the test materials to enable a full critical analysis. It seems likely, however, that any test in which *drawing* is proof of recall is likely to be confounded by factors that do not relate to visual memory (for example, motor skills).

In conclusion, it is argued that spelling errors by themselves cannot produce clear evidence of the role of visual memory, particularly when the error classification scheme specifies more than two basic categories of error. In the absence of a standard classification scheme, most studies of this type depend on idiosyncratic and unique schemes devised by the particular researchers, which raise questions of reliability.

1.4.2 Variations within the population: subtypes of disability

The idea that individuals differ according to their reliance on visual memory for spelling has led a number of authors to identify subtypes of reading and spelling disability based on cognitive subskills. Johnson and Myklebust (1967) first coined the terms "auditory dyslexia" and "visual dyslexia" to differentiate between poor readers on the basis of error type. Broadly speaking, the subtype hypothesis states that those who have poor serial recall skills are poor at processing sequential information and this results in phonetic errors; those with difficulties in holistic or gestalt processing are poor at recognising words as wholes and have an over-reliance on phonetic strategies. This rationale underpins the diagnostic system of the spelling subscale of the British Ability Scales (Elliott, 1983).

Boder (1973) also used error type as a diagnostic measure, based on her own reading and spelling test. This test attempts to take two important factors into account: firstly, it distinguishes between phonologically regular and irregular spellings; secondly, it establishes through a reading test the subject's sight vocabulary, allowing the tester to create a unique spelling test for each subject on the basis of known and unknown words.

These features are important for assessing the respective contributions made by phonological knowledge and visual memory. The spelling of unknown words can be used to evaluate the subjects' spelling strategies (i.e. how well he or she is able to guess a spelling according to the application of sound-spelling rules), and the relationship between reading and spelling of known words can be used to ascertain whether or not those words are stored as "gestalts". In other words, if subjects make phonologically plausible mis-spellings of words they have read correctly, then visual memory can be said to be deficient.

Boder found that 63% of her subjects (children diagnosed as "dyslexic") could be described as *dysphonetic* - showing deficiencies in the auditory channel. 9% were described as *dyseidetic*, with deficiencies in the visual channel, and the remaining 22% were described as "mixed". Dysphonetic dyslexics, unable to construct words on the basis of phonology, attempt to use visual memory alone to spell known words, and make wild guesses for unknown words (such as "lsn" for *listen*). Dyseidetic dyslexics are able to spell unknown words using phonological knowledge (e.g. "dus" for *does* and "bleev" for *believe*), but make many similar errors on known words, suggesting that these words' graphemic representations are not preserved in the internal lexicon.

Treiman (1984) arrived at a similar distinction based on the analysis of spelling errors produced by 46 children aged between 9 and 10. Unlike Boder's sample, these children were randomly selected and fell within the normal range of ability for their age. In this group there was a tendency towards the use of phonological rules; such children were termed "Phoenician" spellers, while those who relied on predominantly visual strategies were termed "Chinese" spellers. Treiman describes the two subtypes as poles of a continuum, based on correlational data showing a consistent pattern across the spelling of regular words, "exception" words and nonwords.

A similar pattern of results was obtained by Weekes (1994), who distinguished between "lexical" (i.e. visual) and "sublexical" (i.e. phonological) readers, and found that subjects in these groups made spelling errors which corresponded closely to their reading strategies.

Although the literature concerning subtypes of spelling disability shows a fairly consistent pattern of performance, there are a number of difficulties associated with it. Firstly, the sampling procedure is too variable to allow the findings to be generalised (Batchelor, Kixmiller and Dean, 1990). By using subjects already classified as "dyslexic", Boder's (1973) study tells us nothing about general reading and spelling strategies; her findings may represent no more than the severity of her subjects' difficulties. Treiman's (1984) sample (46) is too small to be applied throughout the population as a whole (and appears to have been recruited from only one educational establishment). Secondly, categorising individual subjects according to error patterns is fraught with the same difficulties as categorising words, as discussed in the previous section. As with error classification, the categorisation of subtypes is left largely to the judgement of individual researchers.

More comprehensive studies have since been conducted which identify a subgroup of teenage children who have specific spelling difficulties despite normal reading (Newman, Fields and Wright, 1993; Batchelor *et al.*, 1990). In the Newman study, children were assessed at ages 8 and 13, and a subgroup at the latter age was found which had compensated for earlier reading difficulties but was still 2 years behind the chronological age norm for spelling. Out of 368 children, this group numbered only 10, which makes statistical analysis difficult; nevertheless, they appear to have no major cognitive deficits on a limited range of measures. Batchelor *et al.* (1990) examined a larger range of cognitive skills for a sample of 1347 children. Using regression analysis, they found a number of largely verbal subskills which accounted for both reading and spelling. However, they found only one unique predictive variable for spelling: visual sequential memory.

Two problems are associated with this finding. Firstly, no indication is given of the ages of subjects, which makes any findings hard to explain in terms of a developmental model of spelling. Secondly, the visual memory component is part of a more general measure incorporating a number of visual and motor tasks. It is not altogether clear how the authors have isolated "visual sequencing" from the list of variables.

In summary, while these studies suggest that individuals differ according to their ability to use non-phonological information in spelling, they do not provide evidence that "Phoenician"-type spellers have deficits in visual memory, since no explicit measure of visual memory has been used as an independent variable.

1.5 The spelling strategies of the born deaf and visually impaired

The final set of evidence for the role of visual memory in spelling concerns the spelling performance of subjects who have congenital deficits in either the auditory or the visual system. If visual memory plays a major function in spelling, we would expect the born deaf to show no difference in their spelling of phonologically regular and irregular words, and the visually-impaired should rely almost exclusively on phonology (thus being unable to spell words like *yacht*).

Dodd (1980) found that, for deaf spellers, this pattern held true. Compared with hearing controls, 14 year old deaf children made more errors on regular words and fewer errors on irregular ones. Subsequent research has failed to replicate this finding, however; Hanson, Shankweiler and Fischer (1983), and Leybaert (1992), found evidence of phonological sensitivity in the spelling of undergraduate and primary school-age deaf subjects alike.

Burden and Campbell (1994) compared 15 deaf children of 14 years with control groups matched on chronological age and spelling age on picture-spelling and lexical decision tasks. They showed a similar pattern of results to the specifically disabled spellers of Newman *et al.* (1993). Reading had become largely orthographic, yet spelling was still governed by phonological rule-based strategies. In terms of Frith's (1985) model, they had yet to reach the orthographic stage of spelling development.

This finding is somewhat surprising given the expectation that the deaf would spell orthographically throughout their lives, which lends support to the argument that phonological awareness is essential for the initial stages of spelling (Goswami, 1992; Bradley and Bryant, 1980). The suggestion that the deaf subjects used lipreading as a spelling strategy indicates that they spell rather like the hearing subjects in Sterling's (1983) study. However, it meant that the authors were rather more generous in their classification of phonologically accurate spellings (allowing, for example, "skwrl" for *squirrel*).

The difficulty with drawing any firm conclusions from this field of research is that the literacy abilities of deaf spellers are acquired in a profoundly different way from that of hearing spellers. The role of lipreading, for example, means that the overall picture is too confused to apply the findings from this body of literature to the general study of spelling, however interesting these findings may be.

As argued earlier, the experience of reading enables good spellers to construct an internal lexicon of orthographic representations (Goswami, 1992; Ellis, 1984; Ehri, 1980). If one has difficulty in the visual processing of print, due to an oculo-motor disorder, then it should be harder to store such information in visual memory.

That is the hypothesis outlined by Cornelissen *et al.* (1994), who assessed the spelling strategies of a small number (10) of children with "unstable binocular control" and a control group. Both groups had a mean chronological age of 10 and a mean spelling age of 9, and were matched for reading and IQ. The words selected for the spelling task were divided into regular and irregular sets and were varied according to the number of letters. It was found that "unstable" subjects made significantly more phonologically plausible errors than the "stable" group, and this was interpreted as evidence that their visual memories were insufficiently developed as a result of their ocular disorder.

Although the sample size is small in this study, it contains the most thorough error analysis of any of the papers discussed in this review, and thus deserves to be taken seriously. The errors were assessed using a panel of three independent raters to judge errors for phonological plausibility; their ratings were then correlated and the coefficients of their agreements each exceeded 0.9.

1.6 Summary

Models of the cognitive processes involved in spelling make little mention of the role of visual memory. However, there is general agreement that for certain words, e.g. *yacht*, the spellings must be stored in some form of internal lexicon in order for them to be retrieved accurately in the absence of phonological information. However, there is a clear shortage of experimental evidence for any correlation between visual memory and spelling ability. While the evidence from error analysis seems to suggest that phonological rules are more important than visual memory for spelling in general, the neuropsychological literature suggests that there may be a subgroup for whom spelling difficulties can be attributed to visual memory deficits. Lack of reliable measures of visual memory may have been a factor restricting the amount of research in this area; the following chapter addresses the issue of visual memory testing in general.

CHAPTER TWO: VISUAL MEMORY TESTING

2.1 Introduction

In this chapter, the visual processes which were alluded to in Chapter 1 are examined in more depth. The chapter begins with a brief introduction to the recent history of visual memory research. Iconic memory - in which the immediate visual image is stored for less than a second - is discussed initially, followed by a consideration of the process of verbal labelling, which, it is argued, affects all types of post-iconic memory. There follows a review of the ways in which visual memory researchers have attempted to control for this factor.

Then the two main processes of visual memory, which were identified in Chapter 1 - visual recall and visual recognition - are discussed at some length. These sections examine the distinction between short- and long-term recall with respect to visual memory, and also the ways in which verbal labelling affect recall and recognition respectively.

2.2.1 Iconic memory and speech recoding

The first major post-war study of visual memory in its own right was conducted by Sperling (1960, 1967). In these experiments subjects viewed brief displays of rows of letters through a tachistoscope and were asked subsequently to recall a specific row. Typically subjects recalled around 75 per cent of each row. Since they were unaware which row they would be required to recall, Sperling argued that they were therefore able to retain 75 per cent of the array in the immediate post-stimulus memory, for durations of between 250 and 500 milliseconds. This memory store, described as "iconic memory", is likened to the physical after-image of a bright light on the retina (Neisser, 1967).

Coltheart (1980) has distinguished three discrete levels of immediate visual memory. The first is neural persistence; at this level it has been demonstrated that there is rod photoreceptor activity which continues despite the removal of the visual stimulus (Sakitt, 1976). The second level is visible persistence, which allows a visual memory to be retained over an interval of 300 milliseconds (Haber and Standing, 1969). Coltheart (1980) suggests that this memory process enables us to watch the progress of a tennis match from behind a slitted fence while walking alongside it; although the visual information reaching the retina consists of only vertical slits of the scene, the visible persistence of each slit allows us to form a continuous impression. The third level is what Sperling described in his iconic memory experiments; Coltheart argues that for such information to remain in the memory it must undergo some degree of semantic processing.

At this point, it is perhaps necessary to define visual memory as being a process that is distinct from *spatial* memory. The type of information which one would expect to be preserved in visual memory involves features of a stimulus such as colour, shape and direction of movement (Baddeley, 1990). Spatial memory concerns features based on the location of a stimulus - given that relevant information may be auditory in nature, it is worth distinguishing such memory processes from those that are concerned solely with the visual system.

Nevertheless, as Coltheart (1980) suggests, visible persistence is not sufficient to retain a memory for longer than 300 milliseconds. Even the recall of the letters in Sperling's studies is likely to have involved the phonological system in some way. Evidence for this was provided by Sperling (1967) and Conrad (1964) who found that subjects were liable to make "confusion errors" based on the sounds of the letters rather than their visual attributes. For example, if a subject made an error in recalling the letter B, s/he was more likely to offer P as an alternative than, say, F. This led Conrad to suggest that some form of "speech recoding" was operating upon the visual (iconic) trace, since this pattern of errors was similar to that produced using *verbal* presentation of letters. In a later experiment, it was found that the same phenomenon could be observed using pictorial stimuli - a picture of a CAT was more likely to be confused with a RAT than a DOG (Conrad, 1972).

If, even over short intervals, verbal information can exert such a strong influence on visual memory, it is likely to render the experimental study of visual memory fraught with difficulty. As this chapter will explain, this has indeed been the case.

2.2.1.2 The working memory model

In this section, visual memory is discussed in terms of more general models of memory.

The modal model of memory of Atkinson and Shiffrin (1968), devised shortly after Sperling's studies, had little to say about the storage of visual information. Most of the studies on which it was based involved auditory presentation of lists of words, letters and digits, for example Miller's (1956) digit span task. The model assumed three levels of storage; an immediate sensory store, which was modality-specific, like iconic memory; a short-term store with a limited capacity of approximately seven chunks of information; and a long-term store, in which material could only be registered though rehearsal. The more rehearsal it received, the stronger the memory trace would be. Because it derived its evidence partly from the recency effect observed in studies of free recall of word lists (e.g. Glanzer and Cunitz, 1966), it did not attempt to explain how visual information might be stored in the long term.

The modal model was challenged by, among other studies, a single-case study reported by Shallice and Warrington (1970), whose patient KF showed preserved long-term memory despite a digit span of only one chunk. If such an individual was unable to rehearse, it was argued, then surely it must be possible for information to reach long-term memory through some other kind of route (Baddeley, 1990).

As a result, a model of working memory was proposed (Baddeley and Hitch, 1974) in which the short-term memory store was replaced by an active information processing system comprising an attentional control system, the central executive, which is served by two "slave systems", the phonological loop and the visuo-spatial scratch pad. Rather than portraying short-term memory as a passive receptor of information, that simply feeds information into a long-term store, it was argued that a number of sub-processes were necessary which manipulate, transform as well as store, information (Baddeley, 1990).

The concept of these two slave systems is crucial to the study of visual memory, as this chapter will go on to explain. What made them revolutionary within the field of memory research is that they allowed for information to be processed within two modalities - visual and phonological. The phonological loop is the structure which is said to be responsible for verbal rehearsal; it would have a major role to play, for example, in a digit span task. It acts as both an auditory receptor, dealing with auditorily-presented information, and as an articulatory process (rehearsal) (Baddeley, 1990).

Evidence for the phonological loop has been provided by a number of studies. Firstly, the phonological similarity effect reported in section 2.2.1 (e.g., Conrad, 1972), where the phonological similarity of words or even pictures makes recall more difficult owing to confusability. Secondly, it has been demonstrated that recall of long words produces more errors than recall of short words (Baddeley, Thomson and Buchanan, 1975); the determining factor seems to be the amount a subject can articulate in a period of two seconds (Baddeley, 1990). Thirdly, it has been found that unattended speech can affect recall performance even if material is presented visually (Salamé and Baddeley, 1982). Finally, preventing subjects from subvocal rehearsal of items also disrupts recall performance (Baddeley, Lewis and Vallar, 1984). This technique is known as *articulatory suppression* and will be discussed in more detail later in the chapter.

The visuo-spatial scratch pad has inspired much less literature than the phonological loop, partly because its workings have not been fully described (Baddeley, 1986). Some evidence has been provided to suggest that irrelevant presentation of visual material can interfere with verbal recall (Logie, 1986). It has been suggested that it is visual *imagery* that maintains information in this structure, although, as suggested in section 1.3.3, such a hypothesis may be impossible to test. Nevertheless, there is some experimental evidence that the imagery involved in this system may be *spatial* - based on localisation - rather than visual - that is, related to properties such as brightness (Baddeley and Lieberman, 1980).

The central executive has been described as the most important part of the working memory model (Eysenck and Keane, 1995), and it has been suggested that this part of the system is responsible for the allocation of attentional resources (Baddeley, 1990). It is said to co-ordinate the functions of the phonological loop and visuo-spatial scratchpad, making it an attentional system rather than a memory store.

The remainder of this chapter will examine the way in which psychologists have attempted to obtain a separate measure of visual memory.

2.2.2 The role of verbal labelling in visual memory

As discussed in section 2.2.1, even over brief periods subjects are likely to encode visual material verbally (Conrad, 1972). The difficulty faced by visual memory researchers is that, despite presenting material visually, by the time it is recalled or recognised by the subject, it may have been entirely transformed from its internal representation. The evidence for this type of transformation is wide-ranging, but a few key studies will be discussed here.

Probably the first explicit study of verbal labelling was undertaken by Carmichael, Hogan and Walter (1932). Nearly 100 adult subjects took part in an experiment which required them to draw from memory a set of visually presented figures. As each figure was presented, the experimenter read out a verbal description ("*The next figure resembles* ..."(*ibid.*, p.76)). The subjects were divided into two groups which differed only in the list of verbal descriptions heard. Each figure was sufficiently ambiguous to resemble either description; for instance, two circles joined by a short line were described as "eye glasses" in one condition and "dumbbells" in another.

When the two groups' drawings were later analysed, almost a third differed substantially from the original stimuli, with vital segments missing, or new segments added. Less than half were reproduced without being distorted in some way. It was found that most of the distorted drawings resembled the verbal description that their artists had been given. So a subject in Group 1 might have made the straight line curved so as to resemble more closely a pair of glasses, and a subject in group 2 may have made the line thicker and more continuous to resemble dumbbells.

Two explanations could account for this finding. Firstly, the dual mode of presentation gave subjects an option: they could either rely on their visual memories or verbal memories, and a third of them had given preference to the verbal store. Secondly, it could be that the immediate storage process of each subject was a blend of visual and verbal information, with both channels activated simultaneously. The resulting memory was, therefore, a synthesis of both presentation modes.

Further evidence for this phenomenon was provided by Loftus and Palmer's (1974) study of eye-witness testimony, where subjects watched a film presentation of a road accident and were asked questions subsequently about some of the details. Again, subjects proved highly susceptible to verbal descriptions. When asked how fast two cars were travelling "when they smashed", speed estimates were significantly higher than when asked how fast the cars were travelling "when they contacted". Loftus interpreted her findings as evidence for "*the Gestalt hypothesis that progressive memory changes in the direction of a 'better' figure occur autonomously*" (Loftus, Miller and Burns, 1978, p.30).

Although Loftus's work has considerable implications in the legal domain, it could be argued that, at an experimental level, the material she presents her subjects with is too difficult or ambiguous to elicit accurate recall, and that the apparently "autonomous" memory transformations are simply an experimenter effect. This conclusion was also reached by Binet in some of his pioneering work on child witness testimony (reported in Ceci, Leichtman and Bruck, 1995). He concluded that erroneous responses to these types of questions "*reflected gaps in their memories, which they reasonably attempted to fill in order to please the experimenter*" (ibid., p.326).

Much of the subsequent work studying verbal labelling has concentrated on the developmental aspects of the phenomenon. The use of a spontaneous verbal labelling *strategy* seems to emerge during the ages of six to eight (Flavell, Beach and Chinsky, 1966). This type of strategy is dependent on some kind of subvocal rehearsal, described by

Vygotsky (1962) as an internalised version of overt speech, otherwise known as "inner speech" (Gathercole and Hitch, 1993).

The way such a strategy develops was studied in some depth by Hagen and Kingsley (1968) and Hagen, Meacham and Mesibov (1970). In the first of these studies, groups of children aged respectively 6, 8, and 10 years were shown a series of cards with pictures of animals printed on them. The experimenter laid the cards face down and the child had to turn them over in the order that they had been presented. Experimental subjects at each age level were required to label the pictures on presentation by saying out loud the name of the animal. Only the 8 year olds benefited from this procedure. It was argued that the 6 year olds had not yet begun to engage in rehearsal, and that the 10 year olds might prefer to use their own (spontaneous) strategies, which were interfered with by the required overt labelling.

This latter hypothesis was tested in the second study, where a group of college students was also administered the task in addition to the 6, 8 and 10 year olds. The overt labelling condition was also found to be detrimental for this group. The subjects were later asked "*if they had any special way of remembering where the animals were in the row*" (ibid., p.55). The question produced a variety of answers detailing different strategies adopted by the subjects, which were supported by observations of the experimenters, who noticed that the children frequently whispered names to themselves or pointed at previously presented cards while performing the task.

A characteristic of Hagen's studies is that, although for older subjects overt labelling had a detrimental effect *overall*, there was a marked recency effect in the labelling condition. This suggests that, for the final item in the row, recall was facilitated by saying the name out loud. This points to a major methodological weakness in these studies - no consideration was made of the interval between presentation and recall of each item. Hence the final item was presented only a couple of seconds before the subject was asked to recall the order of the row, while the delay between presentation and recall of the first items was considerably greater. While these studies can be said to demonstrate successfully the emergence of labelling strategies, their significance as studies of short-term *memory* is doubtful.

Gathercole and Hitch (1993) explain the development of subvocal rehearsal, or verbal labelling strategies, as synchronous with the development of the phonological loop subsystem of working memory (see previous section). Bjorklund and Coyle (1995) suggest that strategy *use* is distinct from the acquisition of technique. Younger children may be able to acquire the techniques of strategy use but experience "utilisation deficiency" when it comes to putting them into practice. This theory has some similarities with Brown and Loosemore's (1994) theory of spelling development (see section 1.2.1.4), in which developmental dyslexia was said to result from limited computational resources. In that case, the component of working memory most strongly associated with the development of verbal labelling strategy may well be the central executive

2.2.3 Verbal labelling and test materials

Given the evidence for the role of verbal labelling, much visual memory research has been concerned with controlling this interfering variable. One way has been to use testing materials which do not leave themselves open to verbal labelling strategies.

However, this has not proved easy. Vanderplas and Garvin (1959) asked subjects to produce verbal labels for a large number of randomised computer-generated polygons. These shapes varied in complexity, ranging from 4-sided figures to 24-sided figures. Shapes of high complexity proved harder to label than shapes of low complexity, although the labels given for highly complex shapes were more varied and imaginative, which suggests a greater degree of mental effort invested by the subjects.

This interpretation would seem to be borne out by the findings of Kelly and Martin (1974), who used the same stimuli in a recognition task and found no effect of stimulus complexity. Indeed, Clark (1965), in a recognition task, found the *reverse* effect to that of Vanderplas and Garvin (1959): using the same stimuli, he found that subjects were *more* likely to use labels for complex shapes, relying on holistic recall for simple shapes. These findings suggest that subjects will go to considerable lengths to use verbal labels as a memory aid under experimental conditions.

Phillips (Phillips and Baddeley, 1971; Phillips, 1974; Phillips and Christie, 1977) claims to have circumvented this problem in visual memory tasks by using "*abstract patterns designed to be amenable to our visual descriptive capacities but not to our verbal descriptive capacities*" (Phillips and Christie (1977), p.119). The stimuli in these studies were generated by highlighting randomly selected cells in a square matrix presented on a computer, varying in complexity from 4 x 4 squares to 8 x 8. Phillips' results in recognition tasks are impressive, displaying no effect of complexity over delays of 500 milliseconds (Phillips, 1974), perhaps providing support for the concept of iconic memory. Over delays of 3 seconds, performance fell to chance levels for the 8 x 8 matrices, although the 4 x 4 matrices were still recognised at well above chance levels after intervals of 9 seconds. However, little consideration is given to the possibility that the 4 x 4 matrices may have been susceptible to labelling strategies of some sort.

A more parsimonious set of test materials was developed by Cleaves (1977) for a visual recognition task. These consisted of right isosceles triangles, differing only in size and orientation. The goal of the study was "*to optimize conditions for wholistic template image* *processing*" (ibid., p.199) - that is, create stimuli which subjects would not be able to label and would therefore produce a measure of "pure" visual memory. In the study, a target triangle was presented to the subject using a slide projector. This was followed, after an interval of 500 milliseconds, by a test triangle; subjects were required to indicate whether or not the test triangle was identical to the target. It was found that subjects' decision times were significantly faster for identity matches (i.e., true positive responses) than for nonidentity matches (true negative). These results were interpreted as evidence that the stimuli were being processed "wholistically" by subjects, using a visual template strategy. As in the Phillips studies, no clear attempt is made to investigate subjects' strategies using self-report techniques.

It seems that the creation of appropriate stimuli relies heavily on the experimenter's assumption that subjects are not using verbal labels. It may prove impossible to determine this. If so, then other methods need to be adopted to prevent subjects using verbal labels. In the next section, the technique of articulatory suppression is discussed - a method which can be adopted for use with any stimuli, no matter how "verbalisable".

2.2.4 Articulatory suppression

The technique known as articulatory suppression (AS) was first used by Murray (1967, 1968). In studies of short-term memory for lists of letters, subjects were required to say the word "the" at the moment of presentation, to prevent them from the speech recoding effect observed by Sperling (1967) and Conrad (1964). Murray suggested that the additional motor activity would have the effect of suppressing any advantage conferred by using a labelling strategy, thus enabling "*an investigation of STM for sensory inputs relatively uncontaminated by rehearsal*" (Murray, 1968, p. 683). As with the studies by Hagen and

Kingsley (1968) and Hagen *et al.* (1970), marked recency effects were obtained in both studies, which make it difficult to evaluate the precise effectiveness of the technique.

Nevertheless, AS has since become highly popular among visual memory researchers. Various methods have been used for producing irrelevant verbal material during stimulus presentation. In one study, subjects were required to count from 1 to 10 while being presented with sentences (Levy, 1977). In a more recent study investigating the phonemic similarity effect for pictures (Ford and Silber, 1994), children as young as three were required to say 'la la la' during presentation. (Unsurprisingly, their performance on the task was almost at floor level.)

Evidence for the effectiveness of AS is provided by Baddeley, Lewis and Vallar (1984), who found that recall of word lists was significantly affected by AS - but only when AS took place during both presentation *and* recall. This suggests that, for AS to work effectively, it needs to disrupt more than just the initial encoding of the stimulus; it needs to disrupt the rehearsal mechanism that maintains the memory store as well. Other studies which appear to support the effectiveness of AS include Hulme, Silvester, Smith and Muir (1986), and Salamé and Baddeley (1982).

One criticism of AS is that it achieves its effect by using up attentional resources (Parkin, 1988). This possibility was considered by Baddeley, Eldridge and Lewis (1981), who included a non-verbal distractor task (tapping) in a short-term memory experiment. This failed to produce the same task disruption as AS, which was taken by the authors as evidence that AS was causing *selective* disruption of the articulatory rehearsal system.

However, this view has been challenged by Margolin, Griebel and Wolford (1982), who argued that articulatory suppression "*does not appear to achieve its effect exclusively from* *a disruption of phonological recoding*" (*ibid*., p. 617). The tapping task used by Baddeley, Eldridge and Lewis (1981), they claimed, was ineffective compared with AS because it placed less of a load on attentional resources. In their study of recognition for previously presented sentences, they administered a non-painful, "threshold" electric shock to one group of subjects during stimulus presentation. This task was considered similar to AS in terms of its load on attentional resources. It was found that the same pattern of results was produced by the shocks as by AS; evidence, they argued, that AS leads to a general disruption of cognitive processing rather than phonological coding alone (Margolin *et al.*, 1982, p. 616).

While the methodology of the above task is attractively persuasive, it is not in itself a successful refutation of AS. On the surface, it tells us more about their shock treatment than about AS as such (the effects of which had already been well-documented). More convincing evidence comes from the low overall performance levels by AS subjects in the other studies using the technique (e.g., Ford and Silber, 1994).

The other difficulty in accepting the validity of AS is that the precise nature of the interfering task is inconsistent. Each study employing AS appears to use different methods to produce its effect, and clearly these methods vary in difficulty. A numerical task (e.g., counting backwards) is likely to achieve more disruption than a simple motor task (e.g., saying 'la la la'). Furthermore, the actual administration of such a task is highly cumbersome and may result in variation in performance by subjects. Some subjects may enter into the task with gusto, others barely at all (possibly through self-consciousness!). It is for these reasons that, in this project, the design of non-verbalisable stimuli is considered a more effective means of controlling verbal labelling than articulatory suppression.

2.3.1 Visual recall and sequential memory

Earlier in this chapter, it was shown that some short-term visual sequential memory tests may have variations in the delay between presentation and recall of individual items (e.g. Hagen *et al.*, 1970). A similar conclusion was reached by Frick (1985), who argued that, because of recency effects, sequential presentation of stimuli does not actually test short-term *visual* memory. Such testing procedures have been adopted over time as a complement to verbal memory tests for lists of auditorily presented items; however the visual store is quite different and only able to deal with simultaneous presentations of stimuli.

This reflects the findings of Jensen (1962) which were reported in Chapter 1. He argued that simultaneous presentation is more likely to mirror the visual memory processes involved in spelling. Link and Caramazza (1994) argue that spellings are represented spatially rather than ordinally, though as discussed in section 2.2.1, this might suggest a type of memory process which is not exclusively visual. Furthermore, as was discussed in the previous chapter, such a theory has proved difficult to model computationally. However, this theory could be interpreted as suggesting that words are represented as shapes rather than collections of discrete letter units, in which case the simultaneous presentation is perhaps more relevant than a temporal one.

Consequently any references made henceforth in the thesis to "visual sequential memory" will implicitly assume that the information is presented simultaneously. An example of a test of this type is the VSM subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk, McCarthy and Kirk, 1968). In this task, subjects are shown a row of small white plastic tiles on each of which is drawn a simple but unique geometric shape. After a five second delay, the row of tiles is then scrambled and the subject asked to rearrange the tiles in their original order.

A number of other tests have employed a similar format, such as the short-term visual memory subtest of the Aston Index (Thomson and Newton, 1982) and the Immediate Memory for Visual Recall subtest of the British Ability Scales (Elliott, 1983). Frick (1985) has argued that these tests are better regarded as tests of *spatial* memory than sequential memory; however, it could be argued that, in Western society at least, a row of symbols of this type (and, for that matter, a row of letters in a word) tends to be processed sequentially, in a left-right fashion. Hence the skills involved in these tasks differ markedly from those employed in, for example, a block-tapping task (e.g., Milner, 1971), which might also depend on complex saccadic eye movements.

2.3.2 Short-term and long-term recall

The tests described in the previous section have all been measures of visual recall in the *short term*. The distinction between short- and long-term memory has been the subject of a great deal of research, mostly concerned with capacity - the amount of information that can be retained over a given interval. It has also been concerned mainly with retention of *verbal* information that has been presented auditorily. An example of this is the digit span task (e.g., Miller, 1956), where subjects are tested for immediate verbal recall of a string of numbers read out by the tester.

The first detailed model of memory that postulated separate storage systems for short- and long-term memory was constructed by Atkinson and Shiffrin (1968). This model assumes that all information is held initially in a short-term store, from which a selected amount of information is preserved in a long-term store. Although the subject of much debate, the distinction between short- and long-term memory capacity is still a feature of contemporary ideas about memory (see Baddeley, 1990). In terms of time limits, short-term memory is generally regarded as lasting for up to 5 seconds (Baddeley and Scott (1971), while tests of long-term memory usually allow for intervals of 20 seconds or more (e.g., Tzeng, 1973).

For reasons explained in the previous section (with reference to Frick, 1985), it is not so easy to calculate the short-term memory capacity for a sequence of visually-presented items because the material takes longer to present than in the auditory mode. As a result, recency effects become more pronounced. Furthermore, as discussed throughout this chapter, there is the problem in deciding whether material is being held in a truly visual store or, through the use of labels, in a verbal store.

As a result, little research has been conducted into long-term recall for sequential information. The few long-term visual recall tests in use tend to require somewhat different modes of recall from tests like the ITPA VSM subtest discussed earlier. For example, in the Delayed Memory for Visual Recall subtest from the British Ability Scales (Elliott, 1983), subjects are shown a card with line drawings of several familiar objects, and then asked to recall those objects verbally. Hence it cannot be regarded as a true test of visual memory, since an essential requirement is that subjects recode the stimuli into a verbal form.

Other long-term visual recall tests, particularly those used with clinical populations, require subjects to recall stimuli by drawing them. The visual recall subcomponent of the Doors and People test (Baddeley, Emslie and Nimmo-Smith, 1995) is one such test. Subjects copy four patterns (Celtic crosses) and then, after a delayed period, are required to reproduce the patterns from memory. The problem with this type of task is that it involves the deployment of a separate, unrelated skill (motor co-ordination) which may act as a confounding variable.

2.3.3 Visual recognition

As a result of the difficulties in studying long-term visual *recall*, long-term visual memory research has concentrated largely on *recognition*. Another reason is that recognition tasks are easier to administer, particularly in a controlled environment (using computers and so on). Several of the studies reviewed earlier in this chapter have used recognition tasks (e.g., Phillips, 1974; Cleaves, 1977). These tasks frequently take the form of a "yes/no" recognition task as in Cleaves (1977), where subjects are required to indicate simply whether a single test stimulus is the same, or different to a previously presented target stimulus. Or they may follow a "forced choice" format (e.g., Phillips, 1974) where subjects have two or more alternatives to choose from in the test phase.

One question that needs to be answered is whether recognition tasks tap the same cognitive skills as serial recall tasks. Bahrick and Boucher (1968) set out to investigate if subjects' verbal recall of a set of pictures was correlated with their ability to recognise the pictures in a multiple choice recognition task. No correlation could be demonstrated between the two measures; furthermore, no effect of overt labelling at presentation could be discovered. However, the type of information required by subjects for each task was quite different. For the recall condition, subjects had merely to remember the presented items in any order. Hence it cannot be regarded as a test of serial recall. The recognition task appeared to be a good deal harder, subjects having to select the target stimulus - e.g., a cup - from a row of 10 cups very similar in appearance. Therefore the ineffectiveness of (overt) labelling is not surprising.

This finding has led some researchers to comment that "*although there is quite a lot of evidence that explicit labelling of pictures leads to higher recall of the object labels, there is rather little evidence that labelling influences recognition memory for the pictures*" (Humphreys and Bruce, 1989, p. 218). This distinction may only apply to forced-choice tests, where "distractor" items can be presented alongside the target figure. This technique was adopted for the Biber Figure Learning Test (Glosser, Goodglass and Biber, 1989), in which each target item was paired with a distractor which resembled the target sufficiently to hamper the use of a labelling strategy. Thus, if the target design consisted of an inverted triangles embedded in an upright triangle, the distractor consisted of a pair of similar triangles which were not embedded, but adjacent.

Nevertheless, Cleaves (1977) selected single triangles as stimuli because he argued that more than one geometric design may be encoded differently - that is, not as a "wholistic template" or gestalt. It could be argued that this different encoding is likely to involve a verbal label at some stage; the examples provided by Glosser *et al.* (1989) from the Biber test appear to be easy enough to supply verbal labels for. Further evidence that labelling is likely to be involved in recognition comes from a study by Goldstein and Chance (1971), which found that accuracy of recall was correlated with the "meaningfulness" of the stimuli used. "Snowflake patterns" elicited accurate recall of 33%, while inkblots elicited accurate recall of 48%, and faces 71%. This study used a "yes/no alternative" task, which suggests that this format too may be confounded by labelling strategies.

In general, it appears that recognition tests are probably *easier* than recall tests. In a study by Ellis and Daniel (1971) using the Vanderplas and Garvin (1959) stimuli, subjects maintained consistent recognition accuracy over a 28-day period. However, their recall of "meaningful" labels for the shapes showed a significant decline over this interval. Nickerson (1965) found that recognition accuracy for a set of 600 pictures was 92% the following day; even after a year, accuracy was still above chance level (63%).

The relative ease with which subjects perform forced-choice recognition tasks seems to depend on the number of choices. Where the subject only has to indicate one of two alternatives, "*the subject needs only to have stored the minimum amount of information that will allow one of the pictures to appear slightly different from the other. That means* something *on the picture is stored, not that everything has been remembered*" (Baddeley, 1990, p. 21). Therefore, it would appear that the most thorough tests of recognition are those which either employ a "yes/no" technique, or offer the subject several alternative distractors. A recent test of the latter type is the Visual Recognition subcomponent of the Doors and People test (Baddeley, Emslie and Nimmo-Smith, 1995), in which coloured photographs of doors are used as stimuli.

2.4 Summary of Chapter 2

The purpose of this chapter has been to discuss in detail the measures used by researchers to study visual memory, and the potential difficulties involved. It is now hoped to apply the findings of this research in the study of visual memory processes in spelling.

CHAPTER THREE: VISUAL MEMORY, VERBAL LABELLING AND READING

3.1 Introduction

Compared with the small amount of literature concerning spelling and visual memory, there has for some time been an interest in the role of visual memory in the process of reading. The first part of this chapter builds on the research reviewed in the previous chapters and examines the application of visual memory testing to studies of reading ability. The evidence for the role of visual memory in reading is somewhat mixed, so the first two sections compare studies which support the visual memory hypothesis with those which fail to support it. The following two sections review studies which account for the visual memory hypothesis in terms of verbal labelling. The final section discusses the role of verbal labelling in reading as distinct from its role as a confounding variable in measures of visual memory. This section reviews studies of naming speed, using a variety of stimuli, which have been shown to discriminate between good and poor readers.

3.2.1 Studies supporting the visual memory hypothesis in reading

"*[Visual serial ordering] involves the analysis of the order of component symbols in an array; these symbols must be processed according to their visual characteristics since they have no ready name equivalents...this systematic, analytical visual perception skill...must be necessary in grapheme by grapheme reading*" (Ellis and Large, 1988, p. 63).

A number of studies of short-term visual sequential memory over the years have successfully discriminated between good and poor readers (e.g., Rizzo, 1939; Kass, 1963; Hirshoren, 1969; Guthrie and Goldberg, 1972; Crispin, Hamilton and Trickey, 1984; Ellis and Large, 1988; Brannan and Williams, 1988). As the above quote suggests, the cognitive skills required to retain the order of an array of visually presented stimuli have been considered a necessary pre-requisite for the development of reading skill (e.g., Kirk and Kirk, 1971). However, this finding is by no means consistent across the literature (Golden and Steiner, 1969; Holmes and McKeever, 1979; Bell, 1990). This section will consider those studies that appear to support the hypothesis that visual memory is a significant determinant of reading ability.

As discussed in Chapter 2, visual sequential memory (VSM) tests are a common feature of neuropsychological test batteries, which are often used in the diagnosis of reading difficulties. An example of such a test is the VSM subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk, McCarthy and Kirk, 1968), whose basic format was described in section 2.3.1. This test battery, and its authors' remediation programmes, have exerted a "seminal" influence on the field of specific learning difficulties (dyslexia) (Pumphrey and Reason, 1991, p. 92).

The VSM subtest in particular has been found to discriminate between good and poor readers. For example, in a correlational study with the California Achievement Test in second-grade Californian schoolchildren, Hirshoren (1969) found that it correlated more highly with reading (.61) and spelling (.65) than any other subtest of the ITPA. Guthrie and Goldberg (1972) compared two groups - a "normal" reading group and a "disabled" reading group - of 8-9 year-olds on this task. Separate reading-VSM correlations for the two groups showed a significant correlation (.47) for the normal readers, but only a very low correlation (.15) for the disabled readers. Raw scores are not specified in this paper, though Macione (1969) found a significant ($p<.05$) difference between "disabled" and "nondisabled" readers on the same test.

A more recent study of visual sequential memory and reading failed to replicate these results among a sample of British schoolchildren (Crispin, Hamilton and Trickey, 1984). In this study, the ITPA VSM subtest was administered alongside two new VSM tests which used letter strings rather than abstract shapes as stimuli. While one of the new tests was significantly correlated (0.64) with reading age, the ITPA test had a correlation of precisely 0.00 with reading age. Several major reservations must be made concerning this study. The first is the sample size, which numbered only 19. No details are provided about these subjects except that they had a mean reading age of 7.8. Furthermore, no details are supplied concerning the procedural aspects of the experiment, which makes it difficult to comment on the findings.

Two recent studies also make claims for the importance of visual sequential processing in reading. Ellis and Large (1988), in a longitudinal study of reading development, identified "visual serial ordering" as one of the best of many cognitive predictors for early reading development. However, like Crispin *et al.* (1984), they fail to provide a full summary of the task details. Nevertheless, they are explicit about the type of skills required to perform such a task - for example, pattern analysis. This concept is upheld by the findings of Brannan and Williams (1988), who asked groups of "normal" and "disabled" readers to judge which of two stimuli appeared first on either side of a computer screen. Perceptual thresholds for the subjects had been determined beforehand, and it was found that the disabled readers were less accurate at identifying which of the two stimuli appeared first. This was discussed in terms of "perceptual grouping" (ibid., p. 443).

3.2.2 Studies *not* supporting the visual memory hypothesis in reading

Studies using the ITPA VSM subtest have not always found it a clear discriminator of good and poor readers. Golden and Steiner (1969) found no significant difference between groups of normal and disabled readers matched for age and IQ, although the sample size was very small (only 10 subjects in each group).

A more recent study (Bell, 1990) also disputed the ITPA subtest's discriminative validity. In this study, 42 "ordinary" readers were compared with the same number of "dyslexic" subjects (whose reading had been assessed at two years behind their chronological ages). Both groups had a mean age of slightly over 13 years. The groups did not differ significantly either on the number of errors made or the time taken to perform the task. This result could be interpreted as suggesting that, while VSM appears to be a good predictor of reading ability in younger children (Ellis and Large, 1988), its predictive validity is less powerful for older children.

A study that seems to support this view is Holmes and McKeever (1979), which examined VSM in teenage dyslexic subjects (age 13 years). The group sizes were rather small, consisting of 15 "dyslexics" (selected on the basis of attendance at a reading clinic) and 15 controls. There was a slight IQ difference between the groups. Subjects were tested on their ability to recall sequences of 20 words and 20 photographs of human faces, presented individually for 3 seconds at a time on pieces of card. They were then required to reconstruct the order using the cards. Mean scores for the two groups were identical in the faces condition but controls displayed superior recall in the word condition.

This result was interpreted as evidence that dyslexic readers do not suffer from a general VSM impairment. However, the choice of faces as non-verbal stimuli must be challenged. It has been demonstrated that faces possess unique properties that lead to perceptual grouping as early as the first few weeks of life (Wilcox, 1969). The cognitive skills required for memory of faces are likely to be quite different from those used for memory of other stimuli, both verbal and non-verbal (Ellis and Young, 1988). Therefore, faces would seem to be unsuitable stimuli for this type of comparison.

Up to this point, the studies under consideration have all been studies of serial recall. In actual fact one would consider visual *recognition* to be a process more analogous to reading. However, studies of visual recognition have so far failed to discriminate between good and poor readers. One such study, conducted by Ellis (1981), tested the recognition memory of dyslexic subjects and controls using a set of stimuli similar to those used by Phillips (1974) - matrix patterns with different squares shaded randomly. No significant differences could be found between the performance of dyslexics, age-matched controls (12 years), and undergraduate subjects. However, a question must be raised about the reliability of the data in this study, since the numbers of correct responses on the task far exceed those in the original Phillips study, reaching as high as 99 per cent in one condition. It appears therefore that the task proved too easy for all subjects and that a ceiling effect prevented any group differences from reaching significance.

A reservation could also be made about the findings of (A.W.) Ellis, McDougall and Monk (1996), who used a test of "memory span for abstract shapes" as part of a larger battery of cognitive tests comparing the performance of "dyslexic" readers with three control groups. On this measure they found no significant difference between the groups. However, no details are provided of the stimuli (although a source reference is cited) and it does not appear that test design played a large part in the investigation.

Probably the most cited of all studies not supporting the visual memory hypothesis is Vellutino, Steger, DeSetto and Phillips (1975). This again suggested that visual recognition played little part in reading ability. The subjects in the study were drawn from three age groups (7, 10 and 12 years) and were divided into poor readers and normal readers. The stimuli consisted of Hebrew letters, and so subjects were also screened as to their knowledge of that language. The task replicated the findings of an earlier study, in which subjects had been required to draw the Hebrew letters from memory (Vellutino, Steger, DeSotto and Phillips, 1973). That particular study could be criticised on the basis that it tapped a second skill (drawing) which might have interfered with the one under investigation. It was also concerned with "immediate", or short-term recognition, whereas the later study also measured recognition over 24 hours and 6 months. No significant differences could be found between any of the good and poor readers over the three time periods. However, the groups who were familiar with Hebrew letters showed a significant advantage in the immediate condition, and over 24 hours.

There are a number of questions that need to be raised about these results, which show the reverse pattern of the Ellis (1981) study, in that the percentage of correct responses seems uniformly low. Indeed, only the Hebrew group attained 50% correct recognition. This could be a reflection of the task requirements. Unlike most forced-choice recognition tests, where target items are tested individually, in this study the presentation of the target items and the forced-choice recognition phase were administered as discrete blocks. In other words, subjects saw a sequence of 9 Hebrew letters, for 5 seconds at a time, and were then asked to recognise all 9 target figures from groups of 3 alternatives. Therefore, as in the Bahrick and Boucher (1968) study discussed in section 2.3.3, the overall poor performance could be attributed to task difficulty. The fact that even the Hebrew groups recognised only 2 out of 9 items in the 6 month condition suggests that the comparisons are unreliable.

In summary, it appears that the inconsistency of the task requirements and the stimuli employed mean that it is difficult to draw many firm conclusions from the study of visual memory and reading. Indeed, the positive findings from the studies using the ITPA VSM subtest have been challenged by Hicks (1980), and the next section considers this study in depth.

3.3 The Hicks study and related findings

The findings of Hicks (1980) bring together two of the themes of chapter 2 and the present chapter: the role of visual sequential memory in reading, and the confounding variable of verbal labelling.

It is worth beginning by examining some of the claims of the ITPA's creators. In the companion book to the battery, *Psycholinguistic Learning Difficulties* (Kirk and Kirk, 1971), it is made clear that the test's designers were fully aware of the problems posed by potential verbal labelling of the stimuli. "*The task [of visual sequential memory] is sometimes facilitated or circumvented by using mnemonic devices involving meaning or by verbalizing so as to utilise auditory memory*" (p.116). With this in mind, the revised edition of the subtest featured stimuli which had been specially redesigned "*to minimize possibilities of vocalisation*" (p.19).

Hicks tested 20 nine year old children on the ITPA VSM subtest and found that 13 of them were able to label the stimuli on the basis of their resemblance to common objects. (This information was obtained from subjects' self reports). For example, one four-sided figure was frequently labelled "star", while a circle with protruding spokes was labelled "catherine wheel". The difference in test scores between the 13 labellers and the 7 non-labellers was deemed to be significant at the 0.005 level.

In the next experiment, the subjects and task remained the same, with one small modification. This time, subjects were instructed beforehand to assign a verbal label to each stimulus. If a subject was unable to do this, the experimenter supplied an appropriate label. The results showed only a slight improvement by the 13 labellers; however, the non-labellers improved to a significant degree ($p<0.001$). These results suggested that, contrary to the intentions of Kirk and Kirk (1971), verbal labelling is a major factor underlying performance on this test.

A further two experiments were conducted by Hicks to determine whether this factor accounted for the superior performance of good readers. In the first of these, 12 normal readers aged 9 were compared with 12 dyslexics aged 9 with a reading age of 7. This experiment had two conditions. In the first phase, the subjects were administered the test as normal, and then divided into labellers and non-labellers on the basis of self-reports; in the second phase the test was repeated using an articulatory suppression technique (repeating the word "the").

As predicted, the control group showed a preference for labelling; only 1 of the 12 reported using a visual strategy, compared to 9 of the 12 dyslexic subjects. As before, labellers and non-labellers differed significantly in the first condition, but not in the second (AS) condition, where the control group's performance deteriorated to the level of the dyslexic group.

In the final experiment, two new groups of dyslexics and controls were administered the test. This time, all the controls reported using labels and all the dyslexics reported using a visual strategy. In the second condition, all subjects in both groups were asked to label the stimuli. This resulted in a significant increase in performance by the dyslexic group (though the controls were still superior to a significant degree). Hicks interpreted these findings as being clear evidence that "*for too long some aspects of literacy difficulties have been erroneously attributed to visual memory deficits*" (Hicks, 1980, p. 24).

Swanson (1978) arrived at a similar conclusion to Hicks (1980). She tested 2 groups of 30 subjects on a serial recall task using some of the figures from the Vanderplas and Garvin (1959) study discussed in chapter 2. One group consisted of normal readers aged 9, and the other group was selected from special needs classes on the basis of reading disability. Each group was divided into two, depending on whether or not subjects learned labels for the figures in the pre-test phase. The labels were supplied by the experimenter. They were intended to be suggestive of each shape, though the shapes were chosen on the basis of their low "association value". The training phase consisted of repeated exposure to the stimuli; in the "naming" condition, a pretesting requirement was that they could name each figure.

The recall phase employed a "*probe-type serial memory task procedure*" (Swanson, 1978, p. 541). The subject was shown 6 shapes, each exposed "for a few seconds" and then laid face down. Then a target (probe) item was displayed, and the subject was required to indicate which of the 6 cards matched the probe. It was found that only the normal readers benefited from the pretest naming training. The performance of poor readers did not differ significantly between naming and non-naming conditions, and normal readers in the non-naming condition performed at a similar level to poor readers. Thus it seemed that only the good readers benefited from labelling.

On the face of it, this result appears to support the findings of Hicks (1980). However, there is an important difference between the two studies. In Hicks' study, the experimenter only supplied labels where absolutely necessary. In the Swanson study, not only were *all* labels supplied by the experimenter, but they had minimal associative value to the stimuli. Therefore one might have expected similar results to Hagen *et al.* (1970) and H.L.Swanson (1984), in that, with children of age 9, the enforced labelling strategy would interfere with subjects' idiosyncratic labelling strategies. A replication of this study with a slightly older sample might perhaps produce different results. This may explain why, unlike in Hicks' study, poor readers did not benefit from labelling. It seems that it is not labelling *per se* that improves performance, but the instruction to use a labelling strategy.

This distinction may provide an explanation for the findings of a study of short-term visual recall by Torgeson and Goldman (1977). The children in this study were slightly younger (age 8) than in Hicks (1980) or Swanson (1978). They were divided into "good" and "poor" readers on the basis of reading grade level. They were tested using line drawings of 7 common objects arranged in a different order on each page of a booklet. The experimenter pointed to the drawings in a particular order and then, after an interval of 15 seconds, asked the subject to point to the drawings in the same order.

Subjects were observed during the 15-second interval and it was found that good readers were significantly more likely than poor readers to display signs of subvocal rehearsal ("verbalisation"). The indicators for this were lip movements, particularly if these movements corresponded to the object names. Furthermore, good readers achieved significantly higher scores on the test. This finding was taken to demonstrate the importance of mnemonic strategies for reading ability. Further analysis, in the form of subjects' self-reports, suggested that poor readers showed no awareness of rehearsal strategies.

There is an assumption on behalf of some of the researchers of strategy development that age is the main factor determining whether or not subjects use verbal labelling strategies (e.g., Bjorklund and Coyle, 1995; Hagen *et al.*, 1970). The three main studies reported in this section suggest that strategy use may be impaired in some subjects up to 9 years of age, and that this is a reliable predictor of reading ability. Miles (1993) has stated that "*the central difficulty for dyslexics is one of verbal labelling*" (ibid., p. 186).

It appears that, while there is still some evidence for the importance of visual memory in *spelling*, most of the support for the visual memory hypothesis in reading can be accounted for by good readers' superior ability to use verbal labelling strategies. The next section examines the findings of a number of studies which have investigated the role of verbal labelling in reading independently of visual memory. These are mainly studies of naming speed, using a variety of stimuli.

3.4. Studies of naming speed and reading

In the last 25 years, a number of studies have found that good and poor readers differ in the speed with which they can name visual stimuli (e.g., Denckla and Rudel, 1974; Felton, Wood, Brown, Campbell and Harter, 1987; Wolff, Michel and Ovrut, 1990; Badian (1993, 1994); Wilson and Cline, 1995). There are two basic experimental methods that have been employed for this type of study: Rapid Automatised Naming (RAN) tasks, where subjects are required to name stimuli presented in simultaneous sequences; and discrete trial (DT) naming tasks, where stimuli are presented individually. Several studies have compared both methods. The type of stimuli used include letters, numbers, colours and drawings of common objects.

3.4.1 Rapid automatised naming (RAN)

The RAN task was first used in a series of studies by Denckla and Rudel (1974; 1976a; 1976b), who found that poor readers aged between 8 and 10 made more errors, and were slower, than good readers at naming sequences of visually-presented stimuli. This finding held true for letters, numbers, colours and objects. The stimuli were presented on four charts (one for each type of stimulus), each consisting of 5 different items repeated 10 times at random in a 10 x 5 matrix. Naming speed for each chart was calculated by use of a stopwatch.

This format has become widely used by educational psychologists (see Wilson and Cline, 1995). Several recent papers support its predictive power in discriminating between normal readers and dyslexic subjects for colours and objects (Felton *et al.*, 1987; Wolff *et al.*, 1990; Wimmer, 1993; Badian, 1993, 1994), and for digits and letters (Bowers and Swanson, 1991). It has also been found to discriminate between good and poor spellers (Bear and Barone, 1991), but only letters were used in this study. Subjects' age does not appear to be a significant factor - those tested in these studies have ranged from as young as 6 (Badian, 1992) to adults (Wolff *et al.*, 1990). Wolf (1991) argued that, on the basis of a 5-year longitudinal study, that speed at letter and number naming was a good predictor of subsequent reading difficulties.

The major difficulty in interpreting these findings concerns methodological inconsistency. Failure to standardise a RAN task has meant that each study has used original stimuli and mode of presentation. Several researchers have stuck to the basic Denckla and Rudel (1974) format, but others have varied the matrix size to 6 x 8 (Bowers and Swanson, 1991), 8 x 4 (Wimmer, 1993) or even as small as 5 x 5 (Bear and Barone, 1991). On one hand, this inconsistency makes the findings unreliable; yet it could be argued that, since the significant differences are found across such a variety of materials and formats, a typical RAN task taps basic cognitive differences between good and poor readers. Wolff *et al.* (1990) have suggested that the tasks measure speed of access to words in an internal lexicon, and the simultaneous processing of several stimuli.

The idea that poor readers have a deficiency in integrating several sub-skills smoothly is the basis of the "dyslexic automatisation deficit" hypothesis (Nicolson and Fawcett, 1995, 1990). This hypothesis derives from the theory of Shiffrin and Schneider (1977), who argued that performance on multiple tasks (e.g., driving) is transformed through practice into "automatic" processing. A similar conclusion was arrived at by Swanson (1984), who found that the addition of verbal labels to a visual recall task appeared to hinder rather than help dyslexic subjects, leading that author to suggest that the dyslexics experienced particular difficulty integrating verbal and visual information.

In some respects, the RAN task is a close relative of the VSM task, since both tasks require rapid sequential processing. Wolf (1991) argues that this mirrors the process of reading itself: "*rapid scanning, sequencing and processing of serially presented material*" (*ibid.*, p. 127). Although no such study has been undertaken, it could be argued that good readers' verbal labelling advantage on visual memory tasks may be related to their superior ability to retrieve labels at high speed. The argument that RAN is a measure of naming latency (or "lexical access") has, however, been contended by some authors (Perfetti, Finger and Hogaboam, 1978; Stanovich, Freeman and Cunningham, 1983) who argue that this skill can only be assessed by presenting the stimuli in isolation. In the next section, the findings of a number of discrete trial naming studies are discussed.

3.4.2 Discrete trial (DT) studies of naming speed

DT studies have generally proved less successful in differentiating between good and poor readers. However, as with RAN studies, methodological inconsistencies make it difficult to arrive at any firm conclusions regarding the validity of this measure. Perfetti *et al.* (1978) and Stanovich *et al.* (1983) found that DT naming of digits and letters did not correlate with RAN measures, and that no difference in naming latencies could be observed between good and poor readers. Similar results were obtained by Badian (1993), who found that DT performance correlated with IQ but not reading. However, in this study the stimuli were objects taken from the Boston Naming Test (Kaplan, Goodglass and Weintraub, 1982). This is essentially a vocabulary test and is scored in terms of response *accuracy* rather than latency. Felton *et al.* (1987) used the same test and found a significant difference between good and poor readers. Their task was a modified version of the original, in which subjects were given less time to respond.

Wolff *et al.* (1990) also found a difference between normal readers and dyslexics, using a film projection presentation of colours and objects. Here again, though, no attempt was made to measure response latencies; the only data collected concerned accuracy of response. Probably the most reliable DT study so far undertaken is that of Bowers and Swanson (1991) who used computer presentation of stimuli and measured naming latencies by use of a voice-activated relay. They found that this task was successful in discriminating between good and poor readers.

Recent findings by Ellis, McDougall and Monk (1996) suggest that, if anything, discrete-trial picture naming may in fact display the reverse pattern, with poor readers naming pictures more quickly than controls. In their study, "dyslexic" readers were significantly faster at naming computer-presented pictures than three control groups. However, the age and IQ discrepancies between the groups may account for this difference.

Clearly, research using the DT method is a long way from making the sort of progress that has been achieved with RAN tasks. It could be argued that the problem is a procedural one. Even more than RAN, DT studies have varied in terms of the materials, presentation mode and data collection. In a review of the naming speed literature, Wilson and Cline (1995) have argued that DT studies require equipment that makes them difficult to run outside a laboratory setting. Computer technology is, however, now capable of tackling this problem, and in chapter 6 a DT object-naming test will be described that has been produced for use with a microcomputer. It would seem that this type of task might be a better independent measure of verbal labelling ability than the RAN task.

3.5 Summary of the first three chapters

Compared with visual memory and reading, there has been little attention paid to the role played by visual memory in the spelling process. A number of experimental studies have suggested that visual recall and recognition may be significant factors in spelling ability, particularly with regard to irregularly spelled words. However, there is a problem in obtaining a measure of visual memory that is not affected by verbal labelling of the stimuli. It has been shown that studies which claim that good and poor readers differ in their visual memory ability have also been affected by verbal labelling. Where verbal labelling has been studied in isolation, for example in tests of naming speed, it appears to be better than visual memory as a predictor of reading ability.

The experimental work in this thesis sets out to fill some of the gaps in this literature. Firstly, separate tests of visual memory and verbal labelling will be described, along with measures of their reliability and validity. Then a series of studies will be described in which these tests have been administered to groups of 13 year old children along with standard measures of spelling ability. It is hoped to find out if spelling ability correlates with different measures of visual memory (recall, recognition, short-term, long-term) It is also hoped to find out if spelling ability is related to object naming speed.

CHAPTER FOUR: VISUAL SEQUENTIAL MEMORY: TEST CONSTRUCTION AND PILOT STUDIES

4.1 Introduction

In this chapter, the construction and piloting of the first series of tests of short-term visual recall is described. The format of the Visual Sequential Memory (VSM) subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk, McCarthy and Kirk, 1968) was used as the basis for these tests, because of its success as a predictor of reading ability, and because the skills it taps may be involved in the recall of spelling information. The question of verbal labelling was addressed in the test construction, with the result that three tests, each using a different set of stimuli, were devised.

Test 1 (the Kirk test) adapted the stimuli used in the VSM subtest of the ITPA. Test 2 (the Animals test) used stimuli that subjects would find easy to label. Test 3 (the Triangles test) used stimuli that subjects would find difficult to label.

These three tests were piloted twice. In Study 1 undergraduate students were used as subjects, and it was found that the ease with which the stimuli could be labelled predicted subjects' test performance. Furthermore, subjects who reported using a labelling strategy achieved higher scores on the Kirk and Animals test though not on the Triangles test.

In Study 2, the tests were administered to a group of 15-year-old schoolchildren, and the same pattern of results was obtained despite slight modification of some test items. An additional test was used which had the intention of discriminating between subjects on the basis of whether they used a verbal labelling strategy or not. This turned out to be largely unsuccessful.

4.2.1 Choice of basic test format

Most of the tests described in the previous chapters were unique, idiosyncratic tests devised for experimental purposes. It was thought best to use as a model a test which had been standardised and used extensively with relevance to reading and spelling. The VSM subtest of the ITPA seemed to satisfy these criteria, even though it has been the subject of some criticism (e.g. Hicks, 1980). However, this criticism was aimed primarily at the interpretation of the results; this study and others have shown it to be a successful discriminator between good and poor readers (Kass, 1963; Hirshoren, 1969; Guthrie and Goldberg, 1972). Some of the criticisms levelled at the ITPA as a whole (e.g., Dale, 1972) are concerned with its adaptation of Osgood's (1957) three-dimensional communication model, within which visual sequential memory "*is represented at the intersection of the organising process, the visual-motor channel, and the automatic level*" (Kirk and Kirk, 1971). Other criticisms have been the result of several factor analyses of the ITPA, which display inconsistent results, particularly with children below the age of five (Doughtie *et al.*, 1974). However, visual memory emerges as a clear factor in four of the studies (Meyers, 1969), so it would seem that this subtest is perhaps more reliable than some of the others.

Another argument supporting the use of this test as a measure of short-term visual sequential memory is that it seems to tap a skill that might constitute a major factor in spelling. In section 1.3.1 it was argued that visual sequential memory may be required for an individual to recall the ordinal position of letters in a word (Thomson, 1984; Shallice, 1988). In section 2.3.1 it was suggested that the best means of testing visual sequential memory with regard to spelling was to use a simultaneous presentation of a sequence of stimulus figures. This followed the suggestion of Link and Caramazza (1994), that words may be represented as whole units rather than ordinal collections of individual fragments, and of Frick (1985) who argued that a simultaneous presentation of a sequence of items was a more valid way to measure visual recall than a temporally-spaced sequence of separate items.

Although the choice of the VSM subtest of the ITPA might also be criticised on the grounds of its age, it must be added that more recent standard tests of short-term visual memory are strikingly similar. An example is the VSM subtest of the Aston Index (Newton and Thomson, 1982). It was thought, therefore, that it might provide a useful baseline measure against which to compare the efficacy of other stimuli.

4.2.2 Computerisation

One of the aims of the project was to develop materials which might be used by educational psychologists as a diagnostic tool in the assessment of specific learning difficulties (dyslexia). The emergence of the microcomputer (laptop) has enabled administrators to control variables in a way that was not possible with traditional pencil-and-paper tests. Indeed, Pumphrey and Reason (1991) have claimed that "*the microcomputer is to the educational practitioner what the telescope and microscope were to the astronomer and the biologist, with at least one important difference: the possibilities for developing theory and practice that it opens up are even greater*" (ibid., p. 2). In the case of test administration, computerisation enables an administrator to exercise greater control and accuracy of timing, materials, data collection, and relieves researchers of a great deal of paperwork.

The VSM subtest of the ITPA is administered manually, as described in section 2.3.1, which means that the administrator has to operate a stopwatch and manipulate materials at the same time as issuing instructions. Such demands are likely to lead to inconsistencies in display time and data collection. By incorporating these functions into a computer program that can be run on a laptop machine, it was hoped to maximise the ease, efficiency and accuracy of data collection.

4.2.3 Programming and design

The tests were constructed using Authorware Professional, an "object-oriented authoring tool" (Macromedia Inc., 1992). The basic concept of the Authorware package is that users without any knowledge of scripting language can write programs, simply through manipulating icons on a flowchart.

A program was written that maintained the purpose of the original test. In the original, the subject is presented with a visual display on a piece of card consisting of a sequence of abstract shapes. The display is removed from sight after five seconds, and the subject is then required to reconstruct the sequence using plastic tiles, each printed with one of the abstract shapes.

Although it would have been theoretically possible to recreate the original ITPA subtest using Authorware, such a program would require the user to have considerable experience of controlling a mouse. It was expected that the computing expertise of the subjects would be quite varied, and so any tasks would need to be as simple as possible to eradicate this potentially confounding variable.

Therefore, the display phase of the test was the same as in the original, with a sequence of figures in connected boxes displayed for five seconds. Rather than requiring any manipulation of stimuli, however, subjects are required simply to indicate the relative position of a single figure. After a three-second inter-stimulus interval, there follows a display consisting of an empty row of boxes. Beneath the empty row is a single box containing one of the pictures from the previous display. Subjects are instructed beforehand to indicate in which box the target picture had been seen, by directing the cursor to the appropriate empty box and clicking on it. This response in turn brings up the next display. Each of the three tests consists of 15 items.

It could be argued that this modification of the original ITPA VSM has altered the processing requirement of the task. Where before, subjects were required to hold in memory the relative positions of five figures, this new task only requires the subject to recall the position of one. However, since there is no way in which the subject might guess in which position the target figure is likely to appear (this element being randomised), the processing demands are likely to remain the same. Indeed it conforms to the criteria used to design the original test, which stipulated that the task mirror the processes involved in spelling, with stimuli occurring "*in horizontal orientation, in simultaneous presentation, and in close succession*" (Paraskevopolous and Kirk, 1969, p. 46). Instead of calculating the number of figures recalled in the correct positions for each item, a measure was taken of the speed of response.

The Authorware tool enables programmers to record users' responses and perform mathematical calculations. The program registers the number of the box the user clicks in, and whether that box is the correct answer or not. It also calculates the time that elapses before the user clicks the mouse. After the tests have been completed, it calculates the number of correct responses made by the user in each test, and the mean response time for each test. At the start of the program, the user receives four practice items for which responses are not evaluated.

Appendix 1 contains a number of flowcharts which demonstrate the overall design of the program.

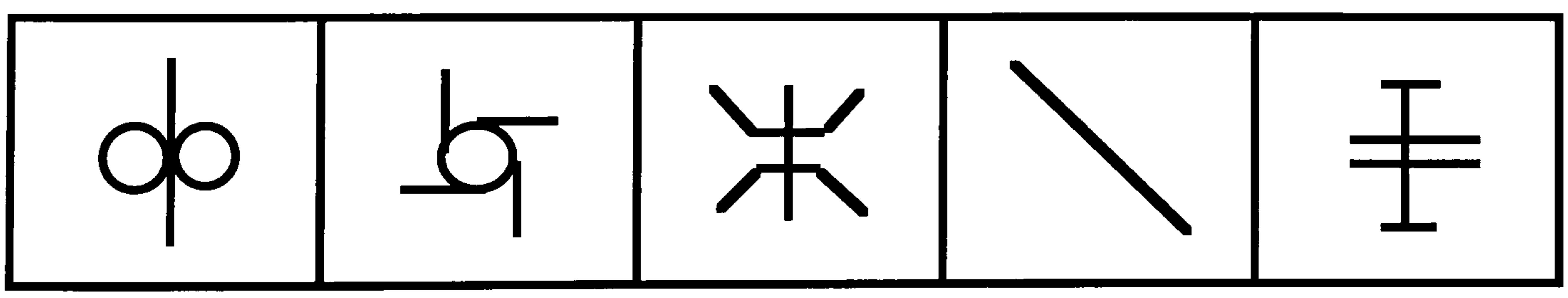
4.2.4.1 Test 1 (the Kirk test)

The first test created for this project was a simple test of short-term visual sequential memory which used three sets of stimuli: a standard set of figures copied from the ITPA VSM subtest, a set of easily labelled pictures, and a set of abstract shapes which were intended to suppress labelling. It was designed to be run on an Apple Macintosh powerbook.

The basic concept of the ITPA task was preserved: subjects were expected to view a row of pictures for five seconds and then, after a delay of three seconds, recall the order of the pictures. In the ITPA subtest, the recall phase requires the subject to reconstruct the original sequence using plastic tiles with the individual pictures on them. However, because of the author's intention to computerise the tests, on the basis of the reasons given in section 4.2.2., it was necessary to modify this phase of the task.

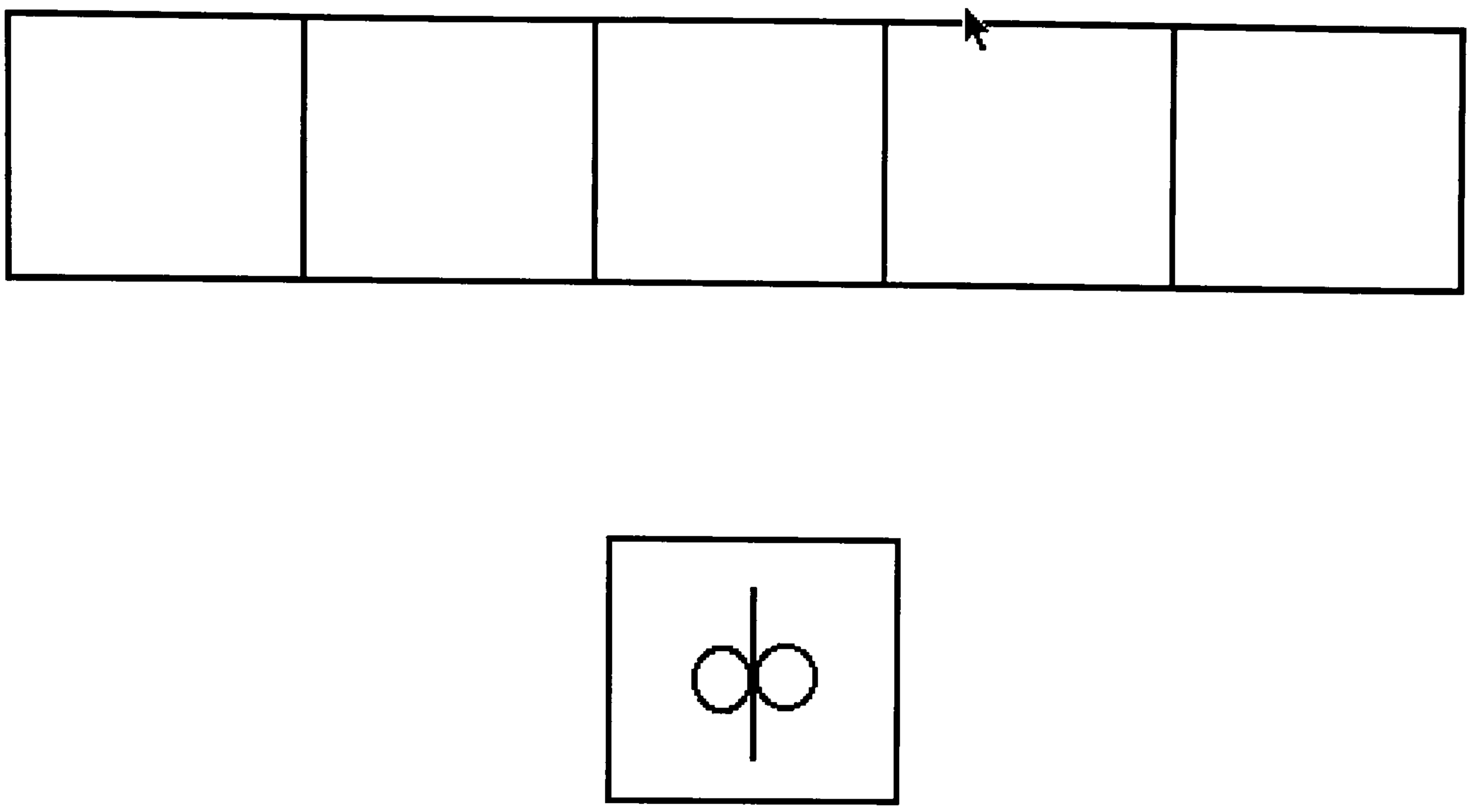
Test 1, henceforth referred to as the Kirk test, uses as its stimuli a selection of symbols taken from the original ITPA VSM subtest. Figures 2 and 3 show the two types of display screen which constitute one item, using some of these symbols.

Figure 2. Detail from the Kirk test: the 5-second presentation of stimuli



Stimuli were selected from the original set on the basis of susceptibility to labelling. Three of the symbols were familiar shapes (a square, a circle, and a diagonal line), while the other four were more complex designs. The full set of stimuli are displayed in Appendix 1 (Figure A2).

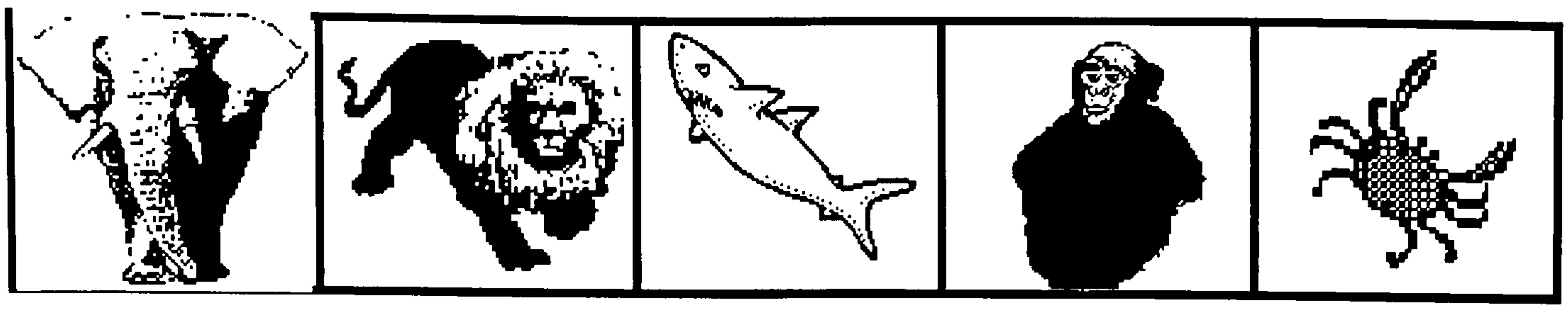
Figure 3. Detail from the Kirk test: the user response screen



4.2.4.2 Test 2 (the Animals test)

For Test 2 (the Animals test) it was necessary to present subjects with stimuli which were so familiar it would be difficult for them to avoid using verbal labels. It was decided to use cartoon pictures of familiar objects which could be copied from "Shareware" applications where copyright would not be a restriction. Figure 4 shows a selection of these pictures. It must be borne in mind that the original stimuli were brightly coloured, and it was only after the first pilot study that the setting was changed to black-and-white. (Again, Appendix 1 contains the full set).

Figure 4: Sample display from the Animals test

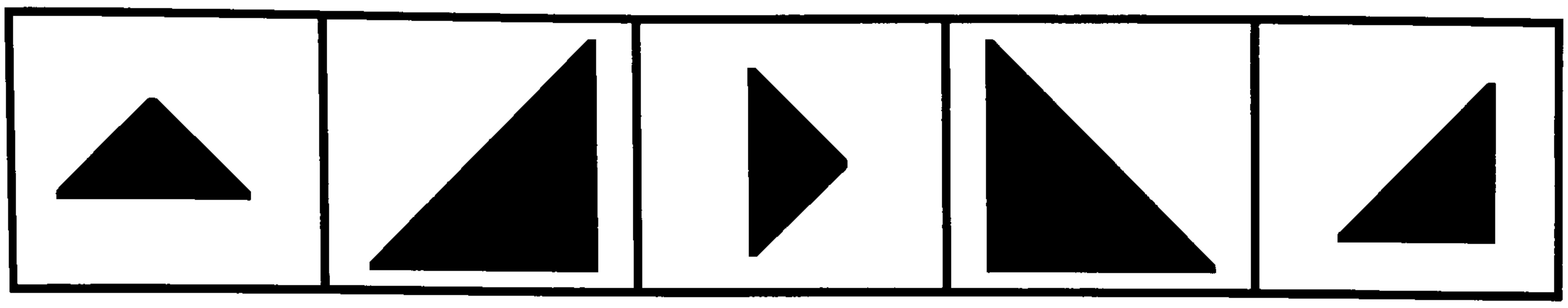


4.2.4.3 Test 3 (the Triangles test)

For Test 3 (the Triangles test), it was necessary to create stimuli which would be difficult to label. As explained in Chapter 2, this has proved a difficult task for visual memory researchers. It was decided that the most successful stimuli for suppressing labelling were either the matrix designs of Phillips (1974) or the triangles of Cleaves (1977). The former have been criticised as "*highly artificial*" by Hitch, Halliday, Dodd and Littler (1989, p. 176). Given that the tests in this project were intended for use with children, it was thought that the stimuli should possess some degree of ecological validity - that is, they should resemble, to some degree, stimuli found in everyday life. It is unlikely that one is ever required to distinguish between patterns so imperceptibly different as Phillips' matrices, without having other meanings attached (e.g., print). As a result, the findings in these studies can be regarded as somewhat "laboratory-bound".

Bearing these factors in mind, a series of triangles was developed based on the illustration in the Cleaves paper of the stimuli used in that study. The triangles in this series differ from one another only in size and orientation.

Figure 5: Sample display from the Triangles test



4.3 Study 1: Kirk/Animals/Triangles pilot

4.3.1 Introduction

The aim of this first study was to administer the tests to a random selection of undergraduate students. This would enable item-total reliability and validity measures to be calculated, and to identify any areas in need of future modification.

It was also hoped to establish the type of strategy subjects were using on all three tests. It was expected that, for the Animals test, subjects would tend to use verbal strategies, while for the Triangles test, visual strategies would be preferred. The Kirk test was expected to show a more equal application of the two strategies.

In the studies of Hicks (1980) and Swanson (1978) it was found that short-term visual memory was most accurate when subjects used a verbal labelling strategy. Therefore it was predicted that the Animals test would yield higher scores than the Triangles test because the stimuli were easier to label. Performance on the Kirk test would, however, depend on which strategy subjects preferred to use. Verbal labellers were expected to achieve higher scores than those who preferred a visual (holistic) strategy.

4.3.2 Subjects

The subjects were 38 undergraduate students aged between 18 and 25.

4.3.3 Method

Subjects were administered the tests individually using a Macintosh powerbook. After completing each of the three tests, subjects were asked to write on a sheet of paper their response to the question: "*how did you remember the symbols in that test?*" This was intended to elicit an answer that could indicate whether or not subjects were using verbal labels in any of the tests.

Ten subjects from the study were also tested on 15 items from the ITPA VSM subtest, administered manually. This was in order to obtain a measure of correlation with the computerised variant, which would indicate how closely the two tasks are related. Although subjects in both tasks need to memorise each item as a complete array, it might be that the computerised variant, where subjects are only expected to identify the position of one figure in the array, might tap slightly different cognitive skills.

4.3.4 Results

Test performance

Mean totals and standard deviations for the three tests are displayed in Table 1. Repeated measures ANOVA found that significant differences existed between the Animals and Triangles tests ($F_{1,37} = 64.8$, $p < 0.01$). This difference is consistent with the initial hypothesis, which stated that the Animals test would be significantly easier because of the facilitative advantage of verbal labelling.

In order to assess the overall validity of the Kirk test, the performance of ten subjects was compared with their performance on 15 manually-administered items from the original ITPA VSM subtest. It was found that the two test scores were positively correlated using Pearson's technique at the level of 0.83 (d.f. = 9, $p < 0.01$), which was taken to suggest that the Kirk test was tapping the same skills as the original. The Kirk test was found to be significantly easier than the ITPA VSM subtest ($t = 3.3$, d.f. $= 9$, $p < 0.01$), though this is not surprising; an item is only scored correct on the ITPA subtest if all the figures in the array have been reassembled accurately.

Table 1. Means, standard deviations, standard error and reliability coefficients for short-term visual memory tests (Study 1)

Test	Mean score (/15)	SD	Standard error	Internal reliability (Gaylord's *r*)
Kirk	**10.7**	2.0	0.66	0.96
Animals	**13.2**	1.6	1.33	0.81
Triangles	**10.0**	2.2	1.54	0.74

Reliability and item analysis

The internal reliability of the three tests was estimated by using the formula given by Gaylord (1969), which calculates a coefficient based on the item-total correlations for each test. These statistics are displayed in Table 1, along with standard error for each test. Full reliability statistics can be found in Appendix 1 (Table A1). Only two items (on the Kirk test and the Triangles test) had individual item-total correlations below 0.5.

A sample of the analysis for the Animals test is featured in Table 2. This analysis was carried out according to the manual of Spurnik and Nuttall (1969) in which the tester compares the performance on each item of the top 25% with that of the bottom 25%. The Facility Index is a measure of overall performance on that item; a high figure (over .75) suggests that the item is too easy - for all subjects - and is thus a poor discriminator of high and low performers, while a low figure (below .25) suggests that the item is too hard and is also a poor discriminator. The Discrimination Index is calculated by dividing high quarter and low quarter scores.

Table 2. Sample of item analysis for the Animals test (n = 38)

Item	High quarter frequencies	Low quarter frequencies	Facility index	Discrimination index
1	9	1	.50	.80
2	10	5	.75	.50
3	10	4	.70	.60
4	10	8	.90	.20
5	9	1	.50	.80

It can be seen that Item 4 had a facility index that was above the upper limit (0.75) specified by Skurnik and Nuttall (1969) for item use, and a discrimination index below the lowest limit (0.25). These figures suggest that the item is too easy and that it needs replacing with a more discriminating item. The other items from this sample fall within the specified range that enables items to be used with confidence. A second item later in the test also proved too easy and failed to discriminate sufficiently between high and low quarters. In the Kirk test there were two items with a discrimination index below 0.21, and in the Triangles there was one item which had a negative value.

In each case the items were subsequently replaced with new items featuring a different arrangement of the stimuli. However, no particular pattern of item facility could be discovered with respect to the position in the array of target figures in any of the three tests. Although there are grounds for only retaining items with high confidence value, and simply eliminating non-discriminating items from the test, for the purpose of means comparison it was felt that the number of items in each test should be identical. Therefore items were replaced with ones which contained the same stimuli as successful items but arranged in a different order. Analysis of these items in Study 2 demonstrated subsequently that they were successful discriminators.

The role of verbal labelling

For each of the tests, respondents were coded as either "verbalisers" or "visualisers". To be allocated to the "verbalisers" group, subjects needed to have made some explicit mention of naming or labelling in their responses when questioned, for example:

"I gave names to the shapes and repeated the list to myself".
"I labelled the items and rehearsed them".
"I named the symbols and repeated the list in my head".
"I named the shapes and memorised them."

"Visualisers" were those whose replies failed to indicate that they had used any kind of verbal strategy, such as:

"I looked at the shapes and tried to remember their position".
"I tried to visually memorise the positions".
"I just looked at them and tried to hold it in my head".

Table 3. "Verbalisers" and "visualisers" in Study 1

Test	"Verbalisers" (*n*)	"Visualisers" (*n*)
Kirk	20	18
Animals	34	4
Triangles	2	36

Table 3 shows the distribution of subjects for each test. From this it can be seen that the majority of subjects chose to adopt a verbal labelling strategy for use on the Animals test, while subjects overwhelmingly chose a visual strategy for use on the Triangles test. McNemar's test of change (McNemar, 1962) found there to be a significant association between test and preferred strategy on the Animals and Triangles tests (c2 = 54, *d.f.* = 1, $p < 0.001$). These figures are in line with the predictions.

A pooled t-test was conducted for the Kirk test, which was the only test in which subjects could be split into roughly equal groups of "verbalisers" and "visualisers". It was found that the verbalisers performed significantly better than the visualisers (t = 3.64, d.f. = 36, p < 0.01). This replicated the findings of Hicks (1980), who found a similar labelling advantage on the ITPA VSM subtest.

On the assumption that the allocation of subjects into verbalisers and visualisers was a general reflection of their preference for strategy use on this type of test, a further set of t-tests was conducted to see if the verbalisers also displayed superiority on any of the other tests. It was found that those subjects identified as "verbalisers" on the Kirk test achieved significantly better scores than the "visualisers" on the Animals test (t = 3.47, d.f. = 36, p < 0.01); their scores were also higher for the Triangles test, though the difference was not significant. Table 4 shows the means for the two groups.

Table 4. The relative means of "Visualisers" and "verbalisers" as identified by the Kirk test (Study 1)

Test	"Verbalisers" (*n*=20)	SD	"Visualisers" (*n*=18)	SD
Kirk	11.7	1.9	9.6 *	1.5
Animals	13.9	1.2	12.3 *	1.6
Triangles	10.6	2.2	9.2	1.4

* significant at the .01 level

4.3.5 Discussion

The reliability statistics for the four tests suggested that the majority of items were discriminating successfully between good and poor performers, and the split-half coefficients and Alpha scores suggested that the tests had good overall reliability. Item analysis revealed only a small selection of items which needed replacing, which were replaced subsequently with suitable items; there appeared to be no particular reason why these items were poor discriminators. There was no consistent pattern of item performance. Order of stimuli did not appear to affect performance; no particular symbol or picture appeared to yield higher rates of accurate recall than any other.

The overall mean scores of the three tests followed the expected pattern. The Animals test yielded higher scores than the Kirk test, which in turn yielded higher scores than the Triangles test, thus demonstrating the superiority of material which lends itself to verbal labelling.

This effect was further strengthened by the way subjects could be allocated to "verbalisers" and "visualisers" groups. Numbers in each group were similar in the Kirk test, but in the Animals test the majority of subjects used explicitly verbal strategies. Conversely, in the Triangles test, only two subjects reported using a labelling strategy. This would seem to suggest that the material used in the Triangles test had successfully counteracted verbal labelling.

The two subjects who used labels in this test had only limited success formulating such a strategy. One subject's response was: *"I started to use labels like "big", and "sideways" but gave up"*. The other subject labelled the triangles according to the direction in which each was pointing ("east", "west" etc.). It would be possible to confound this strategy by ensuring that at least two triangles in each display were pointing the same way.

On the Kirk test, "verbalisers" performed significantly better than "visualisers", which, as in the study by Hicks (1980), emphasises the advantages of using verbal labels for these particular stimuli.

This finding is of particular importance as far as this series of tests is concerned, because it suggests that, although the original ITPA subtest has been modified for computer, it nevertheless taps the same cognitive skills as the original. The demand for subjects to retain the serial order of stimuli remains the same, even though the task requirements are somewhat different. (It could still be argued that the computer test is *easier* than the manual test because subjects do not have to reconstruct all five figures for each item.)

Another issue raised by the superiority of "verbalisers" on the Kirk test concerns the nature of verbal labelling. Apart from their superiority on the test, are there qualitative differences between the two groups that might manifest themselves on other cognitive tasks? This suggestion was tested by analysing the performance of these two groups on the other three tests. Indeed, as Table 4 shows, on the Animals test, the Kirk "verbalisers" also performed significantly better than the Kirk "visualisers". This effect was not deemed to be significant on the other two tests.

A possible explanation for this finding is that, while the Animals and Triangles tests dictate subjects' strategies by the nature of their materials, the figures in the Kirk test are more ambiguous. Subjects *can* use verbal labels, but this is by no means an automatic strategy even for undergraduate subjects. The results would seem to suggest that, where verbal labelling is a subject's preferred strategy, they tend to perform better. Even where the stimuli are easy to label, this group performs better. Done and Miles

(1978) found that the easier it was to label stimuli, the greater the gap between the performance of dyslexic and non-dyslexic subjects. It is possible that the "verbalisers" and "visualisers" in the Kirk test may display different literacy skills. This is a topic which will be returned to in Study 3, where the performance on the Kirk test (and choice of strategy) of good and poor spellers is compared.

It is also apparent that the "visualisers" performed significantly better on the Animals test than on the Triangles test, where their preferred strategy would seem to give them an advantage. To some extent, this finding is in concordance with Hicks (1980) who found that her non-labelling groups (in this case, dyslexic subjects) were substantially aided by using labels. Since most "visualisers" reported using a labelling strategy for the Animals test, it could be argued that, even though labelling is not their "natural" strategy, they benefit from using it. Nevertheless, they still perform at a lower level than those for whom labelling *is* their preferred strategy.

The allocation of subjects into "verbalisers" and "visualisers" on the grounds of self-report must be called into question on the grounds of reliability. It could be that a response such as "I remembered the shapes visually" might actually mask a verbal contribution which the subject had supplied unwittingly. An attempt to tackle this question experimentally will be described in the next study.

One final point concerns the Animals test, in which one subject reported using verbal labels based on the colour of the stimuli. It was therefore decided to alter the setting of the computer program so that the figures appeared in black and white, thus eliminating the possibility that memory for colour names might act as a confounding variable.

4.4 Study 2: Kirk/Animals/Triangles revised pilot

4.4.1 Introduction

In this second study, the same tests were administered as in Study 1. Having piloted the tests initially using undergraduate subjects, it was now decided to use subjects who conformed to the requirements of the project as a whole.

One practical application of the findings of this project relates to the availability of concessions for students with special needs statements at GCSE level. At this stage in spelling development, most children should begin to use "orthographic" spelling strategies (Frith, 1985), for which visual memory should have matured alongside phonological skills (Lennox and Siegel, 1994). Furthermore, the number of irregular spellings children are exposed to should increase during this period, making visual memory more important for spelling overall (Waters *et al.*, 1988). It was thought that if scores on a visual memory test were able to discriminate between good and poor spellers between the ages of 12 (the onset of Frith's "orthographic stage") and 15 (the GCSE age level), that test could have great value for educational psychologists. Such a test might be used to identify whether a child's poor spelling had a genuine cognitive basis, or should be attributed to other factors. It was also felt that, if such a test was capable of identifying poor spellers at age 12-15, then it might be used with younger children to predict potential difficulty in reaching the orthographic stage of spelling development.

Another question under investigation in this study was the reliability of subjects' self-reports concerning their strategies. In Study 1 a significant difference had been found in the Kirk test between subjects who reported using verbal labels and those who did not. Although the results of the Animals and Triangles tests appear to demonstrate the facilitative effect of verbal labelling, this may be because subjects' strategies are dictated by the materials. When the materials are ambiguous - as the shapes in the Kirk test seem to be - subjects may elect to use a verbal strategy or rely on holistic visual processing. So far the only evidence on which to allocate subjects to either visualisers or verbalisers groups has been provided by the subjects' responses to questioning. In this study an attempt was made to test subjects' choice of strategy by experimental means.

4.4.2 Subjects

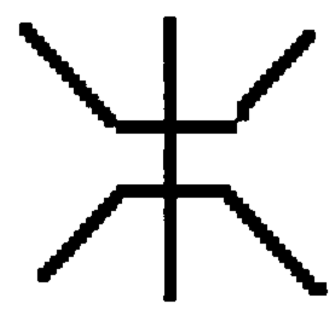
The subjects consisted of 39 15-year-old boys at a secondary school in a West Midlands town.

4.4.3 Method and materials

Subjects were tested on an individual basis, using the same materials as in Study 1, with two modifications.

Firstly, a new task was included with the intention of distinguishing verbalisers from visualisers experimentally, without relying on self-report. At the end of the Triangles test, subjects were then shown a series of ten symbols and asked to supply names for them. Each symbol appeared separately, in a display like the one in Figure 7, and subjects were required to type a name in the space provided. The symbols were all stimuli which had been used in the previous tests. The first five were chosen from the Kirk test, and the second five from the Triangles test. The built-in timer registered the amount of time that elapsed between the onset of each display screen and the subject's first keypress.

Figure 7. Screen dump from "symbol naming" phase of visual sequential memory tests (Study 2)



Can you think of a name for this figure?

Type it here:

Then press the return key and click the continue button

It was expected that subjects whose verbal reports matched the criteria for the "verbalisers" group would take significantly less time to respond to the first five symbols in this test phase. Given that those symbols were the ones used in the Kirk test, and those subjects had deployed a labelling strategy in that test, it was thought that they should have verbal labels readily available for each shape, and therefore should respond more quickly than "visualisers".

Verbalisers were not, however, expected to display any advantage when required to name the next five symbols. Since these symbols were all triangles, differing only in size and orientation, it was thought that all subjects would find them equally difficult to label. In Study 1 it was seen that the vast majority of subjects use largely visual strategies for the Triangles test, and so no subjects would be expected to have quick access to previously-used labels.

The second modification concerned the Animals test. Some concern was felt about the high mean score achieved by subjects in Study 1, and modification was felt necessary. The original pictures had been brightly coloured, and it was suggested that the colour could have been used as a supplementary aid. This suggestion was borne out by the report given by one subject that she "had remembered the colours of the animals". Furthermore, if this was the case, then it could be argued that this was a visual cue, thus detracting from the purpose of the Animals test - to encourage subjects to rely solely on verbal labels (although the subject whose response was quoted may well have used colour *names*). Therefore, the stimuli were converted to greyscale settings so that they appeared black-and-white.

4.4.4 Results

Overall mean scores and reliability data for the three tests are shown in Table 5. Although the overall mean for the Animals test (12.7) is lower than the Animals mean in Study 1 (13.9), it is nevertheless still significantly higher than the score for the Triangles test ($F_{1,38} = 19.9$, $p < 0.01$).

Table 6 shows the segregation of the subjects into verbalisers and visualisers. In the Kirk test, as can be seen, verbalisers accounted for two-thirds of the sample, while for the Animals and Triangles tests the distribution was similar to that in Study 1. Again, McNemar's test of change found there to be a significant association between test and preferred strategy for the Animals and Triangles tests (c2 = 63.4, *d.f.* = 1, $p < 0.001$).

Table 5. Means, standard deviations, standard error and reliability coefficients for visual sequential memory tests (Study 2)

Test	Mean score (/15)	SD	Standard error	Internal reliability (Gaylord's *r*)

Kirk	**10.1**	2.2	0.76	0.91
Animals	**12.7**	2.0	1.16	0.84
Triangles	**10.9**	2.4	1.30	0.83

Table 6. Strategies adopted for all tests (Study 2)

Test	"Verbalisers"	"Visualisers"
Kirk	26	13
Animals	39	0
Triangles	4	35

Table 7 compares the means for all three tests between those designated as verbalisers and visualisers on the basis of the Kirk test. As in Study 1, verbalisers performed better on the Kirk ($t = 2.1$, d.f. = 37, $p < 0.05$) and Animals ($t = 3.6$, d.f. = 37, $p < 0.05$) tests, and also on the Triangles test, although this does not reach significance ($t = 0.8$, d.f. = 37, $p > 0.05$).

Table 7. The relative means of "Visualisers" and "verbalisers" as identified by the Kirk test (Study 2)

Test	"Verbalisers"	SD	"Visualisers"	SD

Kirk	**10.6**	1.9	**9.1** *	2.5
Animals	**13.5**	1.2	**11.3** **	2.6
Triangles	**11.2**	2.4	**10.5**	2.5

* significant at the .05 level

** significant at the .01 level

On the symbol naming task, the Kirk test symbols took significantly longer on average to name than the Triangles test symbols ($t = 4.2$, $p < 0.01$), a finding which goes in the opposite direction than would have been expected. The differences between the two groups (verbalisers and visualisers) did not reach significance.

4.4.5 Discussion

Many of the findings of Study 1 were replicated in this study. The Animals test again proved easiest for most subjects; again, significantly easier than the Triangles test. However, the modification (from colour to black-and-white) appears to have been successful in eradicating the ceiling effect.

The segregation of the sample into verbalisers and visualisers followed a similar pattern to Study 1, except that the number preferring verbal labelling for the Kirk test was greater in Study 2.

Again, a few subjects reported using labelling strategies for the Triangles test, which suggests that it is not entirely free from labelling strategies. In fact, it could be argued that many subjects use some form of subvocalisation to perform the test without actually labelling the items as such. For example: "*Now, that one was pointing that way, and* that *one was pointing down...*"

As in Study 1, those identified as "verbalisers" by their responses relating to the Kirk test achieved higher mean scores on that test than "visualisers". But the difference between these two groups was greater for the Animals test. Even where *all* subjects adopted a labelling strategy for familiar pictures, those who had previously adopted such a strategy for abstract shapes performed better.

However, the question of how best to allocate subjects to "verbalisers" and "visualisers" groups was not answered by the symbol naming test introduced in this study. It had been expected that verbalisers would respond quicker to naming the Kirk test symbols, but there would be no difference between the groups on the Triangles test symbols. In fact the mean response times did not differ between verbalisers and visualisers for either set of symbols. It was also expected that *all* subjects would name the Kirk test symbols quicker than the Triangles test symbols. However, the results went in the opposite direction, with the Triangles test symbol times being significantly lower.

Why should this task have proved difficult for verbalisers? It has been mentioned earlier that labelling is not necessarily a conscious strategy, and many subjects might arguably use labels automatically without being aware that it *is* a strategy as such. Some of the verbalisers displayed considerable reticence when asked to name the symbols, asking questions like: "*Will "squiggly thing" do*?" It seemed that they felt the task required them to type a formal name (as though it were a Rorschach-type association task) and they were embarrassed to type the name that they had previously used as a label.

This is another example of the complex nature of subvocal rehearsal, whereby "labelling" is perhaps not a sufficient concept to describe the way in which subjects use "inner speech" to aid their task performance. Indeed, even the term "inner speech" may not be adequate to describe the way in which subvocal rehearsal is used. Baddeley and

Wilson (1985) describe a case study, GB, who was found to exhibit sensitivity to word length in recall of visually presented material despite suffering from anarthria (dysfunction of the motor output of the speech system). Similar findings have been reported by Vallar and Cappa (1987) and Bishop and Robson (1989), who found evidence of word length effects in children who had been anarthric from birth. These results have been interpreted as evidence that there may be "*a common innate mapping of sounds onto articulatory gestures at an abstract level*" (Gathercole and Hitch, 1993, p.205). Determining subjects' strategy use may prove impossible except where - as in the Animals and Triangles tests - strategies are dictated by the nature of the stimuli.

4.5 Summary

In Chapter 4, three tests of visual sequential memory were piloted on two separate samples. The results of studies 1 and 2 suggest that visual memory *per se* was measured most accurately by the Triangles test, because verbal labelling did not appear to be a useful strategy, and thus subjects were perhaps relying on holistic visual processing, retaining a Gestalt representation of the visual array. The Animals test, with its easily-labelled stimuli, was so strongly influenced by verbal labelling that in some respects it ceased to be processed visually at all, many subjects preferring to recode the stimuli as a list of verbal items. Strategies were mixed on the Kirk test. However, the subjects who achieved the highest scores on the Kirk test were those who used verbal labels. In both studies, the Animals test yielded higher mean scores than the Triangles test.

CHAPTER FIVE: VISUAL SEQUENTIAL MEMORY AND SPELLING ABILITY

5.1 Introduction

This chapter contains a description of the first major study of the project, in which the tests piloted in studies 1 and 2 were used as dependent measures in a study comparing a group of 24 poor spellers aged 14 years with a group of 24 age-matched controls. The poor spellers were selected on the basis of Special Needs teachers' identification of children whose spelling weaknesses could not be attributed to anything other than cognitive factors. However it was discovered that these children were mostly of low IQ, and so the statistical analysis was carried out using IQ as a covariate.

The Animals test was the only test of the three which favoured controls to a significant degree. It was suggested that this might be a reflection of poor spellers' inability to utilise strategies to gain an advantage using easily-labelled stimuli. It was also suggested that poor spellers have difficulty integrating visual and phonological information. The Triangles test failed to differentiate the two groups, leading to the suggestion that poor visual sequential memory may be an insufficient factor by itself to account for spelling disability.

5.2.1 Introduction to Study 3

Having successfully piloted three tests of visual sequential memory in studies 1 and 2 it was decided to administer these tests to a group of poor spellers in order to compare their performance with that of a control group.

On the basis of the three literature review chapters in the thesis, it was expected that the tests would produce the following results.

- The Kirk test would favour the control subjects, because they are more likely to use verbal labelling strategies. Verbal labellers enjoyed an advantage on this test in both studies 1 and 2, and in this respect it was expected that the pattern of results might follow that of the Hicks (1980) study, in which dyslexic subjects failed to use verbal labelling strategies, and performed at a significantly lower level than controls. Whether differences between the two groups attained significance, however, was likely to depend on the distribution of "verbalisers" and "visualisers" across the groups.

- The Animals test was likely to favour the control subjects to a significant degree. This prediction was based on the findings of Hicks (1980) and of the pilot studies, which suggest that verbal labelling is a preferred strategy for the majority of subjects on this task. However, it was expected that poor spellers may experience "utilisation deficiency" (Bjorklund and Coyle, 1995) when it comes to deploying such a strategy. Brown and Loosemore (1994) argue that the errors of dyslexic spellers show evidence of limited processing ("computational") resources, and reduced processing capacity has been shown to have a delaying effect on strategy use (Guttentag, 1995). Lennox and Siegel (1994) have argued that good spellers integrate visual and phonological information more successfully than poor spellers. As a result, one would expect controls to be more adept at the processing skills required in the Animals test.

- The Triangles test, as the strongest measure of visual memory *per se*, was unlikely to yield significant differences between the two groups. Although it appears to tap the skills required for spelling that are suggested by Link and Caramazza (1994) and Shallice (1988), it seems that performance by normal subjects on the task is handicapped by the absence of any phonological information, which renders the strategy of verbal labelling redundant. Therefore, large differences between controls and poor spellers are unlikely.

Differences *within* the poor spellers group may well be expected on the Animals and Triangles tests, given the subtype theories of Boder (1973) and Treiman (1984). These theories argue that it is possible to distinguish between poor spellers with a deficit in visual memory, who spell most words on the basis of phonology ('Phoenicians'), and poor spellers with a phonological deficit, who rely on visual memory to spell most words ('Chinese'). According to this set of theories, Phoenician poor spellers should perform significantly better than Chinese poor spellers on the Animals test (which has a strong phonological component) but Chinese spellers should perform better on the Triangles test (a measure of visual memory *per se*).

5.2.2 Subjects

The subjects consisted of two groups of schoolchildren.

The control subjects were 24 children from a Year 10 class in a West Midlands secondary school (2 additional cases were dropped on the premise that English was not their first language). The group had a mean age of 14:8 years, consisting of 11 boys and 13 girls.

The experimental group, henceforth referred to as the "Poor Spellers" group, was recruited from six secondary schools in a Midlands town. Special Needs teachers were contacted to identify children on the basis of the following criteria:

- they should be drawn from Year 10 (i.e. about 14 years of age).

- their spelling ability should be at least two years behind their chronological age.

- their spelling difficulties cannot be attributed to any of the following factors: general low intelligence, behavioural problems, or educational withdrawal.

24 subjects were recruited by this method. Their mean age was 14 years and 3 months. The group consisted of 16 boys and 8 girls.

Subjects' spelling ability was determined by use of the age-graded word lists from the Boder Tests of Reading-Spelling Patterns (Boder and Jarrico, 1982). These lists were chosen because one of the tests' diagnostic purposes is to classify dyslexic subjects into either "dyseidetic" (visual weaknesses), "dysphonetic" (phonological weaknesses), or "mixed" subgroups. To identify these respective weaknesses, subjects are given lists of words which are classified as alternately "phonetic" or "nonphonetic". Each subject received 20 words from the appropriate list. The word lists can be seen in Appendix 1A.

A measure of non-verbal IQ - the Cognitive Abilities test - assessed shortly before the subjects had left primary school, was obtained from the local education authority for the subjects in the Poor Spellers group. The mean IQ score for this group was found to be 85. This was disappointing in the light of the sampling procedure, where Special Needs teachers were asked to identify poor spellers of average IQ. It suggests that teaching staff may not necessarily be reliable in their (informal) assessment of children's intelligence levels. IQ scores for the controls - again, derived using the Cognitive Abilities test - were supplied by the school. The mean IQ score for this group was found to be 104. The difference in IQ scores between the two groups was found to be significant ($t = 8.5$, d.f. = 46, $p < 0.01$).

Table 8 displays the mean age and IQ scores for the two groups, along with their mean total out of 20 on the Boder word lists.

Table 8. Mean age, IQ and mean score on the Boder word lists for the two groups (Study 3)

Groups (n = 24)	Age	SD	Spelling mean (/20)	SD	IQ	SD
Poor spellers	**14:8**	0.5	**2.56**	2.4	**85**	13.1
Controls	**14:3**	0.3	**14.91**	2.1	**104**	5.0

5.2.3 Procedure

All subjects received the same tests as the subjects in Study 2, although the symbol naming task was omitted since it had failed to identify verbalisers in the previous sample. Also, a measure was taken of the time subjects took to respond to items and a mean response time was calculated for each subject on each test. This was obtained using an arithmetical calculation function which was part of the Authorware programming software. Hirshoren (1969) found that poor readers took longer to reconstruct the sequence of symbols of the ITPA VSM subtest than good readers; therefore a measure of user response time might produce a second criterion on which to evaluate task performance.

Subjects were tested individually. Each child was excused from class for a 20 minute period, and taken to a small office in the school where testing was able to take place without any interruptions or noise disturbance. The visual memory tests were administered after a brief explanation and the practice items, and the three tests were followed by the Boder word lists.

5.3 Results

Test performance

As discussed in section 5.2.2, there was a significant difference between the two groups in terms of IQ. Given this disparity, it was decided to use IQ as a covariate for all the analyses in this study. Table 9 shows the adjusted mean scores on the three visual memory tests for the two groups.

Table 9: Adjusted means of good and poor spellers on short-term visual memory tests (Study 3)

	Kirk		Animals		Triangles	
	Mean (/15)	SD	Mean (/15)	SD	Mean (/15)	SD
Good spellers	**10.9**	2.4	**12.5**	1.6	**10.8**	2.2
Poor spellers	**8.8**	2.5	**10.0**	2.5	**9.5**	1.9

With IQ as a co-variant, significant differences were found between both groups only on the Animals test ($F_{1,47} = 6.6$, $p < 0.05$). The group means did not differ significantly on either the Kirk ($F_{1,47} = 0.4$, $p > 0.05$) or Triangles ($F_{1,47} = 0.6$, $p > 0.05$) tests. On each test the controls achieved higher mean scores than the poor spellers.

As in studies 1 and 2, subjects were allocated to "verbalisers" and "visualisers" groups following completion of the Kirk test on the basis of their answers to the question: "*How did you remember the symbols in that test*?". Table 10 shows the distribution of verbalisers and visualisers for each of the two groups.

Table 10. "Verbalisers" and "visualisers" in Study 3

Group (n = 24)	"Verbalisers"	"Visualisers"
Poor spellers	6	18
Controls	13	11

A chi-square test found this distribution to be significant (c2 = 4.26, *d.f.* = 1, $p < 0.05$). As predicted, the majority of poor spellers fell into the "visualisers" group since they did not make any mention of verbal strategies in their responses. The distribution of verbalisers and visualisers in the control group followed the same pattern as within the pilot samples. Within the controls, as would be expected from the results of studies 1 and 2, the verbalisers achieved significantly higher scores than the visualisers on the Kirk test ($F_{1,23}$ = 7.3, $p < 0.05$). However, within the sample as a whole, the difference in scores between verbalisers and visualisers did not reach significance ($F_{1,47}$ = 3.0, $p > 0.05$).

For each subject, a measure was made of the time taken to respond to each item which had been answered correctly. Mean response times were then calculated for each test and these means were compared between and within the two groups (see Appendix 5). An analysis of variance, again using IQ as a co-variate, failed to find significant differences between controls and poor spellers on the Kirk ($F_{1,47}$ = 0.8, $p > 0.05$), Animals ($F_{1,47}$ = 1.3, $p > 0.05$), and Triangles ($F_{1,47}$ = 0.9, $p > 0.05$) tests.

Subgroups

Spelling errors were analysed with the intention of creating two subgroups of poor spellers on the basis of error type. These subgroups were termed "Phoenician" and "Chinese" following the classification scheme of Treiman (1984). "Phoenician" spellers were those whose errors were deemed to be acceptable phonetic approximations to the target word spellings, described by Boder and Jarrico (1982) as *Good Fonetic* (sic) *Equivalents* (GFEs). "Chinese" spellers were those whose answers were *not* GFEs. Although a number of subjects satisfied these criteria precisely, the majority were allocated to either subgroup according to the greater incidence of error type. For example, if a subject made 12 errors, and 7 were GFEs, then s/he would be classed a "Phoenician". This classification system is slightly different to the one proposed by Boder (1973), in that she identifies a "mixed" group for borderline subjects. Table 11 displays the errors on three selected words of two "Phoenician" and two "Chinese" poor spellers.

In order to validate the allocation of each subject to one of the two subgroups, two independent judges with thorough knowledge of the classification system were asked to perform the allocation task. This follows the procedure adopted by Cornelissen *et al.* (1994). The judges' classification was then compared with that of the author, and all three ratings showed a high level of consistency (r = 0.77 and 0.79 respectively). Both these figures were deemed significant ($p < 0.01$). 16 subjects were classified as "Phoenician" spellers and 8 subjects as "Chinese" subjects.

Table 11. Comparison of "Phoenician" and "Chinese" spelling errors for three selected words from the Boder lists (Study 3)

	PURSUIT	MORTGAGE	CIRCUIT
Phoenician			
AC	persute	morgage	curcit
DS	persuit	morgage	serkit
Chinese			
CC	buldy	meng	cuter
HW	pursit	mortage	circut

It is possible that these subgroups might differ from one another on other measures besides spelling error type. It may be that the Chinese spellers represent the bottom half of the group with regard to spelling ability, or that the subgroups may differ in IQ. This did indeed prove to be the case. Phoenician spellers spelled a significantly higher number of words correctly ($t = 3.9$, d.f. = 22, $p < 0.01$), and were found to have significantly higher IQ scores ($t = 2.7$, d.f. = 16, $p < 0.05$) than Chinese spellers.

Again it was felt that the disparity in IQ scores necessitated the use of a covariate analysis. Using IQ as a covariant, it was found that there were no significant differences between the mean scores of Phoenician and Chinese poor spellers on the Kirk ($F_{1,23} = 0.5$, $p > 0.05$), Animals ($F_{1,23} = 4.8$, $p > 0.05$), or Triangles ($F_{1,23} = 1.5$, $p > 0.05$). However, the results of the Animals and Triangles tests followed the predicted direction; Phoenicians achieved a higher mean score than Chinese on the Animals test, while the Chinese achieved a higher mean score than the Phoenicians on the Triangles test. Adjusted mean scores for these groups can be found in Appendix 5.

Case Studies

Although the group comparisons of Phoenician and Chinese spellers failed to uncover significant differences between these groups of poor spellers, there are a small number of interesting single cases within the poor spellers group whose specific deficits the Animals and Triangles tests appear to have identified. Their mean scores on the three tests are displayed in Table 12.

Table 12. Performance of selected Phoenician and Chinese poor spellers on the Kirk, Animals and Triangles tests (Study 3)

	Kirk	Animals	Triangles
Phoenician			
AC	8	12	9
RB	13	13	7
Chinese			
CC	6	4	11
HW	4	8	6

Subjects AC and RB display a pattern of performance that is not unlike that of the pilot subjects in studies 1 and 2, and the control group in the present study. Their highest scores are on the Animals test, as would be expected from the pilot results. Both score slightly below the poor spellers' mean on the Triangles test. In both cases their spelling errors were mostly GFEs; indeed RB demonstrated considerable ingenuity with his GFEs, for example writing 'hybonation' for *hibernation*. RB appears to be an individual with highly developed phonological skills, and this is reflected in his high score on the Animals test.

The "Chinese" subjects CC and HW produced slightly different patterns of performance from other subjects. HW appears to have a generally poor visual memory as measured by the three tests, and her phonological skills are not sufficiently advanced for her to compensate, so that few of her errors are GFEs. CC achieved a very low Animals score but an above-average Triangles score. Her spelling performance indicates very underdeveloped phonological skills. A typical error is 'hedne' for *hibernation*. She spelled several of the initial phonemes correctly but appeared to have great difficulty with words of more than two syllables. It could be argued that CC has a processing deficit at the level of the phonological loop (Baddeley, 1990), and that this could be attributed to a reduced rehearsal capacity; this might explain her poor performance on the Animals test.

5.4 Discussion

The results of Study 3 followed the predictions made in section 5.1.

The Kirk test, as predicted, favoured the controls, but - using IQ as a covariate - not to a significant degree. It was suggested, on the basis of earlier studies, that performance on this task would be determined by the strategy employed by subjects. It was expected, following the findings of Hicks (1980) and Swanson (1978) that controls were more likely to employ a verbal labelling strategy than poor spellers. This was indeed the case. However, throughout the sample as a whole, verbalisers did not achieve significantly higher scores than visualisers, although within the control group, verbalisers *did* achieve significantly higher scores.

This finding can be interpreted as indicating that, even where poor spellers elect to use a verbal labelling strategy, it does not achieve the same performance-enhancing effect that it does with good spellers. This can be explained by Bjorklund and Coyle's (1995) theory of "utilisation deficiency", and would seem to concur with Brown and Loosemore's (1994) theory of dyslexia "*in which dyslexics eventually have the same processing strategies available to them as normal subjects, but are delayed in their acquisition of these strategies*" (*ibid*., p. 330).

Further evidence to support this theory is provided by the results of the Animals test. In section 5.1, it was predicted that this test would be the most likely of the three to differentiate significantly between controls and poor spellers. Even when IQ was used as a covariant, the difference in mean scores was deemed to be significant, with controls achieving higher scores than poor spellers.

In Chapter 4, it was argued that the results of studies 1 and 2 clearly indicated that most subjects use a verbal labelling strategy on the Animals test. It was therefore predicted that poor spellers would be at a disadvantage on this test because of their apparent "utilisation deficiency". This appeared to be the case in study 3. There are several other explanations for this finding. The Animals test requires a greater amount of processing than the other two tests because subjects are expected to integrate visual and phonological information; presentation is visual but successful performance seems to require phonological processing and therefore "recoding". Since poor spellers are believed to have a reduced processing capacity (Brown and Loosemore, 1994) and are less successful at integrating visual and phonological information (Lennox and Siegel, 1994; Goswami, 1992), they would be expected to experience a disadvantage in this type of task.

If visual memory alone were a sufficient discriminator of spelling ability, then the Triangles task would be expected to significantly favour good spellers. However, as predicted, the difference between the scores of poor spellers and controls was not significant when using IQ as a covariate. Given that the Triangles test appears to tap the visual sequential memory skills which may be employed in spelling (Thomson, 1984; Shallice, 1988), this is perhaps a surprising result. It might indicate that visual sequential memory is not sufficient *by itself* as a spelling route.

Analysis of individual case studies might suggest, however, that scores on the Animals and Triangles tests together might indicate the type of strategies that individual uses in spelling. For example, the subject RB, described in section 5.2, performed above average on the Animals task and below average on the Triangles task. His spelling was characterised by an over-reliance on phonological information, which is what might be expected from his test scores. Conversely, CC, a very poor speller with a clearly limited processing capacity, performed above average on the Triangles test and well below average on the Animals test. This subject appeared to have good visual skills but very poor sound-spelling skills. Thus on tasks requiring phonological information, she was at a considerable disadvantage.

In general terms, the allocation of poor spellers into "Phoenician" and "Chinese" subgroups was not successful because error type appeared to correlate with overall spelling ability and IQ. Typically, the "Chinese" poor spellers were severely disabled spellers of below average IQ. However the difference between subgroups on the Animals and Triangles tests was predicted in section 5.1. Phoenicians achieved higher scores than Chinese on the Animals test, in which phonological information might play a role, but lower scores on the Triangles test, a task requiring holistic visual processing. These differences did not, however, attain significance, and so little can be gleaned from this set of results.

The segregation of poor spellers into two groups may in itself be impractical. Even Boder's (1973) subtype theory specifies three groups, the third being a "mixed" group with deficits in visual *and* phonological processing; it could be argued that a subject such as HW (see 5.2) belongs in this category. One major concern is the poor reliability of Boder's classification system, which relies on subjective interpretation of the incidence of GFEs (Willows and Scott, 1994; Gerber and Hall, 1987). Although there was high inter-rater agreement for the subgroups in Study 3, this could be attributed to the fact that the Chinese spellers were simply the poorest spellers in the group.

One other variable which was measured in this study was response time. This measure was used because it has been found to discriminate between good and poor readers (Hirshoren, 1969), and it was felt that it might offer a secondary level of discrimination if raw test scores were unable to differentiate between poor spellers and controls. However, any differences between the two groups in response time were not significant. This might indicate one way in which these tests differ from the original ITPA subtest, in which completion times are more relevant because the task requires subjects to reassemble all the figures in the array, rather than merely indicating the position of one target.

5.4.1 The use of IQ as a covariant

It could be argued that the contribution of intelligence to spelling, assessed by a measure of IQ obtained by a standardised test battery, is not sufficiently profound to warrant the use of a covariate analysis of the data in the present study. The use of IQ as a valid psychological construct has been the topic of fierce debate throughout the last century (see Gould, 1981 for a detailed critique). IQ scores can be derived from any one of a large number of batteries which consist of different subtests, any one of which might be reinterpreted using the same type of analysis as in this chapter's discussion of the VSM subtest of the ITPA. Therefore, to discuss IQ as an invariant universal concept is somewhat controversial.

A major concern of many critics of IQ is that the tests used may be culturally specific, and that performance may be mediated by familiarity with the stimuli, and indeed by familiarity with the testing situation itself. Some studies have demonstrated the beneficial effect of repeated practice in IQ testing (see Shackleton and Fletcher, 1984).

Even if these criticisms are invalid, and IQ were assessed in every case using the same instrument, it is still not clear how much relevance such a measure would have to spelling. Nevertheless, the aim of this project is to study poor spellers whose difficulties can *only* be attributed to cognitive functioning and not to external factors such as behavioural difficulties, home environment, and general poor cognitive ability, which *ought* to manifest itself in a low IQ score. Many potentially interfering factors - such as the nature of early spelling instruction - cannot be adequately controlled in a study of 13 year old children, so it is necessary to control wherever possible, and where IQ information is available it would seem sensible to control for this in sample selection.

In a future study of this type, where IQ information is not available, and a control of general ability is required, it may be preferable to obtain marks for a range of school subjects to use as a screening instrument. For example, if a poor speller has low marks in English, Maths, History and Science then that child would be excluded from the sample; if a poor speller only attains a low score in English then it would appear that their general intelligence was of an average level.

5.5 Summary and concluding remarks

In this chapter, the results of Study 3 are described. This compared 14 year old poor spellers with a control group on the tests constructed and piloted in chapter 4. Once the variable of IQ had been accounted for, only the Animals test was found to favour the controls to a significant degree. This finding was interpreted as evidence that short-term visual sequential memory is not related to spelling by itself. In this respect it compares with the studies of visual memory and dyslexia (e.g. Hicks, 1980) where it was found that it was verbal labelling which differentiated good and poor readers rather than visual memory *per se*. It was suggested that these findings indicate that poor spellers may be handicapped by an inability to use strategies successfully because they have difficulty integrating visual and phonological information.

CHAPTER SIX: VISUAL RECOGNITION AND NAMING SPEED: TEST CONSTRUCTION AND PILOT STUDIES

6.1 General introduction to Chapter 6

In this chapter, the construction and piloting of two further tests is described. The first of these is the Windows test, a test of long-term visual recognition memory in which subjects are required to perform a yes/no recognition task several minutes after viewing a target set of items. There follows an account of Study 4, which was a pilot study of the Windows test using an undergraduate sample, and a discussion of the item analysis and subsequent modifications.

In the second part of the chapter the construction of the Pictures test is outlined. This is a discrete-trial picture naming task designed to assess the contribution of verbal labelling independently of visual memory. The construction of the test is described, followed by a pilot study (Study 5), in which a number of items were administered to younger children in order to deselect items which might be ambiguous (i.e. where pictures have competing names).

6.1.1 Theoretical basis for the Windows and Pictures tests

In Chapter 5, it was found that a test of visual sequential memory was not sufficient to discriminate between good and poor spellers. It was suggested that this indicated that visual sequential memory alone may not be a plausible spelling route, and may be merely one of a set of visual memory processes required for spelling. In Chapter 1 it was argued that visual recognition may be an important requirement for spelling competence. Therefore it was decided to create a test that would tap this skill.

Since the only test in Study 3 which discriminated between good and poor spellers - the Animals test - was one in which phonological information (verbal labelling) played an important role, it was also decided to create a test that would tap this skill as an isolated process. The results of the Animals test were discussed in terms of strategy use, and it was suggested that poor spellers may have performed badly on this task because of a "utilisation deficiency" - slow acquisition of effective strategy use. Such an interpretation is consistent with the dyslexic automatisation deficit hypothesis of Nicolson and Fawcett (1995, 1990), as described in section 3.4.1. This theory suggests that dyslexic children find it harder to combine subskills in order to achieve an automatic level of performance. If the Animals task could be said to utilise two subskills - visual and phonological processing - then a dyslexic subject is likely to find it difficult.

In terms of Baddeley's (1990) model, then, the key deficit in spelling disability may occur not in terms of the separate operation of the phonological loop or visuo-spatial scratchpad, but in the central executive. This component of working memory could be regarded, in terms of Brown and Loosemore's (1994) phrase, as "computational resources", where integration of visual and phonological information takes place.

An alternative explanation is that the verbal labelling strategy which appears to favour good over poor spellers may be a by-product of some other form of cognitive factor which distinguishes between spellers. In section 3.4, a number of studies were reviewed in which good readers (and spellers) achieved faster times than poor readers and spellers on tests of rapid naming (e.g., Denckla and Rudel, 1974; Bear and Barone, 1991). It may be that it is not *strategy use* in which poor spellers are deficient, but the speed with which they can retrieve words from the internal lexicon. This may be why poor readers and spellers are less likely to opt for a labelling strategy (Hicks, 1980).

It was decided, then, to test this alternative hypothesis by creating separate measures of visual and phonological processing skill. The Windows test was created to act as a supplement to the Triangles test, both tests measuring the visual memory processes likely to be combined in spelling using the orthographic route. The Pictures test was created to test the verbal labelling element (speed of lexical retrieval) of the Animals test in a task where visual memory would not play a role.

The construction and piloting of these two tests is now described in detail.

6.2 The Windows test

In Chapter 1, it was argued that the visual memory processes involved in spelling can be split into two broad areas: visual sequential memory and visual recognition. Study 3 examined the contribution that visual sequential memory might make towards spelling. It was found that, when verbal labelling was *not* a facilitative strategy, as in the Triangles test, visual sequential memory did not appear to be a factor that differentiated good and poor spellers. It may be that a few individuals (such as RB in Study 3) have a deficit at this stage of the spelling procedure.

In Chapter 1 (section 1.3.2) a number of studies were reviewed which suggested that visual recognition might be important for competence in spelling. The ability to recognise mis-spellings would seem to be a basic literary requirement, and the findings of Ormrod (1985) and Tenney (1980) suggested that poor spellers find it harder to spot spelling mistakes when words are printed in an unusual manner. The recognition process might be important at an earlier stage; Sperling (1983) suggests that good spellers use an internal "lexical monitor" to select the appropriate spelling.

It was therefore decided to design a recognition test which could give a clear measure of subjects' ability to retain the configuration of a design over a longer time interval. This measure could then be used as a variable in a comparison study of good and poor spellers.

6.2.1 Choice of basic test format

In section 2.3.3 there was a discussion of the various ways in which recognition could be measured. To summarise, recognition tasks employ either a "yes/no" format - where subjects have to say whether or not an individual item appeared in a previous list - or a "forced choice" format, where the target item is chosen from several alternatives ("distractor items"). A recent example of such a test is the visual recognition component of the Doors & People test (Baddeley et al., 1995). Here, subjects are shown a series of coloured photographs of doors, and then asked to recognise each target door from a set of four alternatives.

However, the forced choice format was deemed unsuitable for use in this particular project because it was not considered the closest analogy to the type of recognition process which is involved in spelling. In a formal spelling test, subjects are rarely in a position to select a correct spelling from a series of alternatives. The "yes/no" test format would therefore seem to be a closer approximation to the process of recognition in spelling. Therefore the basic format would be the same as in the visual recognition component of the Doors and People test, with an inspection phase and a recognition phase.

An important consideration was the length of time that should elapse between these two phases of the test. Few authors have attempted to cite a temporal cut-off point distinguishing short-term and long-term memory, although in a study by Tzeng (1973) it was found that 20 seconds of distracting activity (counting backwards) was sufficient to eliminate the recency effect associated with short-term recall. However, it was decided to use a longer interval, since this would be a measure of memory that would be more relevant to the process of spelling, where words remain in the internal lexicon for many years. An interval of five minutes was decided upon, during which time it would (in the later studies) be possible to administer another test from the battery, which would act as a distracting activity.

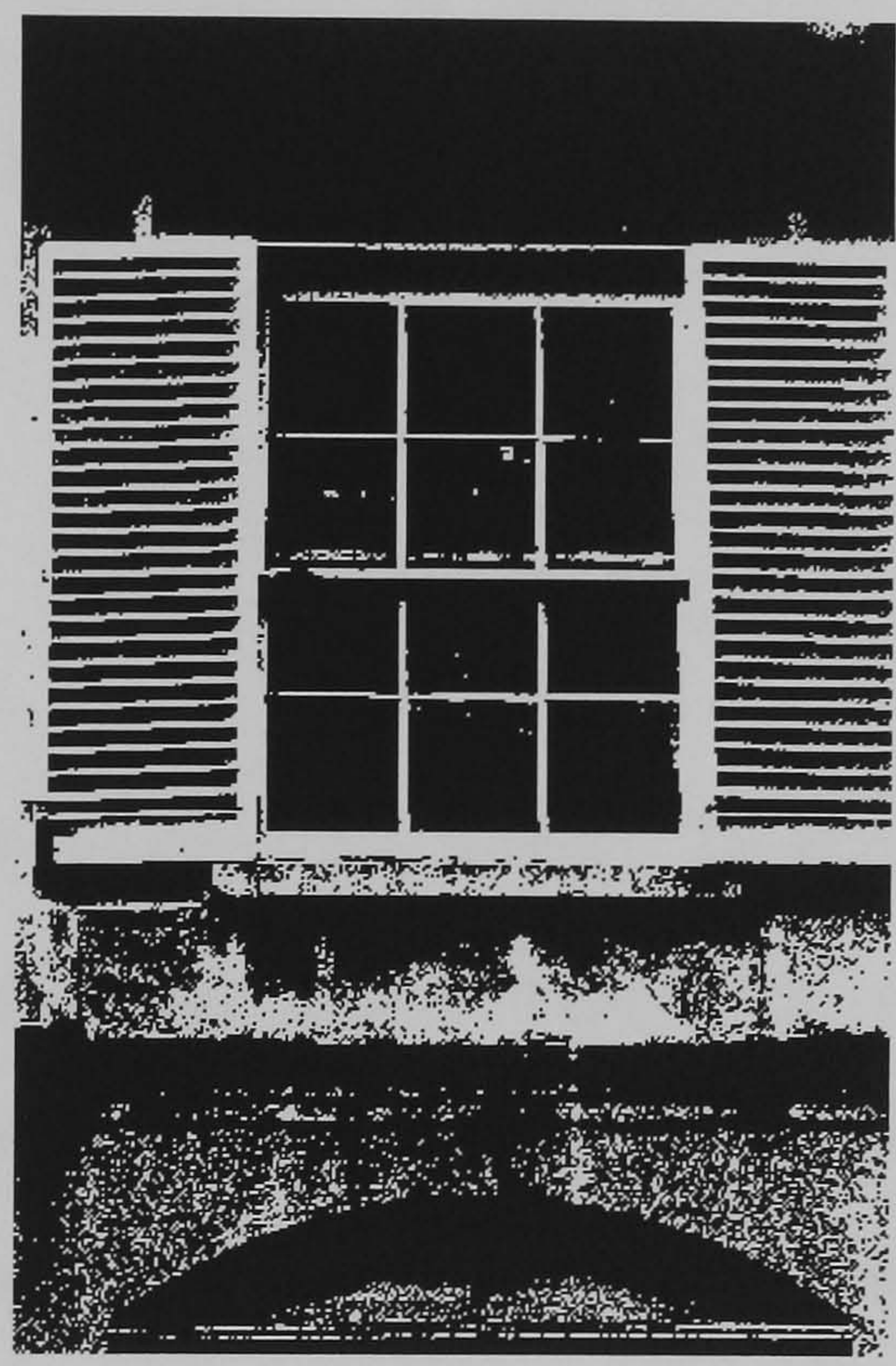
6.2.2 Programming and choice of stimuli

While the recognition component of the Doors & People test was not ideal in its present format for the purposes of this project, the stimuli it employs were thought to be suitable. Photographs of doors were chosen as "*material that is ecologically plausible and readily acceptable to patients*" (Baddeley et al., 1995), and were regarded as stimuli that would be difficult to label verbally.

It was decided therefore to design stimuli that satisfied the objectives of the Baddeley test without infringing copyright. To this end, a number of coloured photographs of *windows* were obtained from houses in a West Midlands town. These stimuli share the same properties as Baddeley's doors; they are objects that are immediately familiar, all share basic features (frames, glass, walls), though each possesses enough unique visual information to be acceptable as a recognition test item.

The photographs were electronically scanned into PICT files which could then be copied into the test program. The test was constructed using the "experiment generator" SuperLab (version 1.68) (Abboud, 1993). This package enables experimenters to design tests without needing any formal programming knowledge. When copied on to a laptop computer, the colour setting of the program was changed from colour to grayscale, to sharpen the pictures and to control for the potentially confounding variable of colour. A similar modification had proved successful for the Animals test, helping to eliminate a ceiling effect (see section 4.4.3). Two examples of the stimuli are shown in Figure 8.

Figure 8. Detail from the Windows Test



A number of trials were carried out using the whole set of stimuli to decide which pictures would make suitable test items. During these trials it was found that subjects were using verbal labels to memorise some of the items. For example, a window might have a black balcony rail, or white shutters, or the branch of a tree might be visible in the left hand corner. Therefore it was necessary to construct the test carefully, so that verbal labellers would not be given the advantage that they enjoyed in the earlier tests.

Where only *one* window in the set possessed a unique feature (for example, a window which had clearly visible stickers on the glass), it was eliminated from the set. Where *two* windows shared the same distinctive feature, it was decided to use them as a matched pair, with one as a target and the other as a distractor. If a single member of the target set had a black balcony rail, then a single member of the distractor set would also have a black balcony rail. A 12-pane target window would have a corresponding 12-pane distractor window, and so on.

A final total of 32 items was considered suitable for piloting; 16 target items and 16 distractors. The full set of windows can be seen in Appendix 2 (Figure A2).

6.3 Study 4: Windows pilot

6.3.1 Introduction

The purpose of this study was to evaluate the 32 items selected for the first version of the Windows test, and to establish a baseline level of performance. As in Study 1, an undergraduate sample would serve as a representative sample of the general population, and would enable reliability measures to be made and item analysis to be undertaken. It was expected that the test would eventually be narrowed down to a smaller number of items.

It was also hoped to compare the test performance of visualisers and verbalisers in order to ascertain whether the "shared feature" distractor items were suppressing the verbalisers' labelling strategies. It was predicted that no difference would be observed between the two subgroups.

6.3.2 Subjects

The subjects were 30 undergraduate students aged between 18 and 25.

6.3.3 Method

Each subject was seated in front of the computer and told that they were about to see 16 photographs of windows from local houses. They were asked to memorise the windows because they would be asked to identify them in the test phase of the experiment. They were also informed that each window would appear on screen for five seconds. This information was provided in case subjects lost concentration towards the end of the inspection phase.

The program was run and, after viewing the target set, subjects were engaged in a distracting activity for 5 minutes before embarking on the test phase. At this point they were seated in front of the computer again and told that they were about to see some more photographs of windows. Although the number of windows was not specified, they were told that some of the windows would be the same ones they had seen earlier and that some were new. They were instructed to press the key marked "1" if they had seen the window earlier, "2" if it was a new window. They were also informed that there was no time limit in this phase of the test, and that each window remained on screen until they pressed a key, which would in turn bring up a new window.

The subjects were left to respond to the full 32 items. Having completed the test, they were asked what technique they had used to memorise the windows, and their responses were coded, as in Studies 1 to 3, so that it would be possible to categorise subjects as either "verbalisers" or "visualisers".

6.3.4 Results & Discussion

Table 13 contains the descriptive statistics concerning the Windows pilot. As in Study 1, internal reliability was calculated using Gaylord's (1969) formula.

Table 13. Descriptive and reliability statistics for Study 4

Subjects (n)	30	Standard deviation	2.73
Items	32	Internal Reliability	0.96
Mean score	22.3	Standard error	0.55

Item analysis

As in Study 1, this analysis was performed following the procedure specified by Spurnik & Nuttall (1969). Table 14 is a sample of the analysis, featuring the final five test items.

Table 14. Sample of item analysis for the Windows test (n = 30)

Item	High quarter frequencies	Low quarter frequencies	Facility index	Discrimination index
28	7	4	.69	.38
29	8	8	1.00	0.00
30	7	3	.63	.50
31	8	4	.75	.50
32	6	2	.50	.50

It can be seen from the above table that, although items 28, 30, 31 and 32 all satisfy the requirements for reliable test items (see Spurnik & Nuttall, 1969), item 29 fails to discriminate between good and poor performers.

Altogether, four items were found to have a discrimination index of below 0.21. It appeared that these items could have lent themselves to verbal labelling. For instance, item 29 was found to be the only window with clouds reflected in the glass. All four were removed from the test subsequently, thus bringing the total number of items down to 28.

On the basis of subjects' answers to the question: "How did you remember the pictures of the windows?" subjects were allocated to either the "verbalisers" or the "visualisers" group. The relative means of these two subgroups are presented in Table 15. No significant difference could be found between their scores ($t = 0.4$, d.f. = 28, $p > 0.05$).

Table 15. Mean scores for "verbalisers" and "visualisers" (Study 4)

	"Verbalisers"			"Visualisers"		
	No.	Mean score(/32)	SD	No.	Mean score(/32)	SD
Windows test	11	**22.3**	3.1	19	**23.3**	2.9

This result appears to support Humphreys & Bruce (1989), who argued that verbal labelling was not a facilitator with regard to visual recognition. However, it must be borne in mind that the test was carefully designed to prevent verbal labellers enjoying an advantage, and that this factor may have affected the result. Indeed, "visualisers" have a slightly higher mean score, which suggests that the "shared feature distractors" strategy has been successful.

6.4 The Pictures test

As explained in section 6.1.1, the Pictures test was created as an attempt to measure the process of "verbal labelling" which might account for the different pattern of results found on the Animals and Triangles tests in Study 3. One explanation was that poor spellers are less adept than good spellers at using strategies; that this was a reflection of limited computational resources; and that poor spellers have difficulty automatising separate subskills. An alternative explanation could be that it is the *speed* of lexical retrieval that differentiates good and poor spellers. In order to test this hypothesis fully, it is necessary to create a measure of verbal labelling which did not feature a memory component.

Numerous studies (see section 3.4) have been undertaken which seem to indicate that poor readers and controls can be differentiated on the basis of naming speed. "*[Rate of access to verbal information in long-term memory is]the most likely candidate for a basic processing difference between reading disabled and normal children*" (Torgesen, 1985, p. 351). Most of these tasks involved naming a number of items presented in a matrix format, and calculating the time taken to name all the items in order. These are known as Rapid Automatised Naming (RAN) tasks. The items may be letters or digits, or the names of colours or objects. Bear and Barone (1991) argued that there is a relationship between naming speed and "orthographic knowledge", and that slow naming speed is characteristic of spellers who have yet to reach the orthographic stage of spelling, regardless of maturational factors.

In devising a measure of naming speed, two major factors needed to be considered:

- *Type of stimuli*

  Most RAN studies have used letters and digits as stimuli, including the only study in which spelling ability was used as an independent variable (Bear and Barone, 1991).

However, it has been argued that naming speed of letters and digits ceases to differentiate good and poor readers beyond primary school age (Walsh, Price and Gillingham, 1988). However, it has been claimed that naming speed for colours and objects is "*developmentally invariant*" (Wolff, Michel and Ovrut, 1990, p. 559), and that this ability differentiates good and poor readers at later ages. Therefore it would seem that object names would be more appropriate a measure to use with the 13-15 age group. The naming speed of objects would also be related to the skills required in the Animals test, so if speed of lexical retrieval were the crucial factor in the Animals test, one would expect poor spellers to display longer naming latencies than controls.

- *Mode of presentation*

  The RAN task is the most commonly used naming speed task; however, as reported in section 3.4.2, a number of studies have employed the discrete trial (DT) format in which naming speed for individually-presented items is measured. Stanovich, Freeman and Cunningham (1983) argued that naming speed in DT presentation of letters and digits was *not* correlated with reading ability; Wolff *et al.* (1990) found that naming speed in DT presentation of colours and objects *was* correlated with reading. However, as with Felton *et al.* (1987), who produced similar results, subjects were not assessed according to naming latencies but to the number of naming *errors* produced.

It was decided to employ two measures of naming speed in the next stage of the project. One would be a traditional RAN test, manually administered; the other would be an original discrete-trial picture naming task. The following sections describe the construction of the Pictures test, and a pilot study carried out to select appropriate items (Study 5).

6.4. Programming and choice of stimuli

Studies of discrete trial picture naming have been carried out intermittently for many years. Perhaps the earliest is Oldfield & Wingfield (1965), who employed complex and ingenious methods in order to obtain a measure of naming speed for pictures projected on to a screen. Response latencies were calculated by how quickly subjects wrote down the object name ("pen reaction"), and "lamp intensity" was used as the independent variable. In thirty years the methodology has not changed greatly; Wolff et al. (1990) used as a measure the number of errors subjects made in a similar task, while Felton et al. (1987) and Badian (1993) used books of pictures and a stopwatch. Only Bowers & Swanson (1991) have attempted to computerise such a task; however, they did not use pictures as stimuli, and, as Wilson & Cline (1995) state, "*the equipment and procedures for the discrete trial format would be much more difficult to reproduce outside a laboratory setting*" (*ibid.*, p. 40). The advent of the microcomputer, though, and the development of "experiment generator" programming packages, has made such criticisms invalid.

Therefore, it was decided to construct a test in which the subject would be required to supply a name for a picture on a computer screen, and a measure would be made of the time taken to supply that name. As with the Windows test, the experiment generating package SuperLab (version 1.68) was used to construct the Pictures test. Although the basic presentation of pictures could be handled by any experiment generator, the important feature of SuperLab is its ability to accept input from an internal microphone. This can be connected to the timing device, which can be activated by voice onset, which makes it an ideal instrument for designing a test of naming speed.

The task itself is simple: present the user with a pictured object (e.g. an apple) and wait for him or her to articulate the name of that object. The voice response is registered by the microphone and a response time recorded in the results file. The closing of the timer is followed by a brief pause before the next item is projected on to the screen, and the test continues in this fashion until all items have been presented.

Pictures of objects, which were likely to be familiar to 13 year old subjects, were found in the Clip Art file provided with the presentation package Powerpoint (Microsoft, 1993). These pictures were able to be copied into PICT files and then, like the Windows, inserted into the program. Unlike the Windows and the Animals stimuli, it was not thought that *colour* would be an interfering variable, and so the images remained in their original format.

Before conducting the pilot study, 50 proposed items (and the test program) were tested using a number of adult subjects. One purpose of this trial was to ensure that the objects were instantly identifiable from the pictures. If adults were uncertain about the precise nature of the object pictured then it would clearly be unsuitable as a test item. Using this criterion, a number of these items were omitted from the pilot study.

A total of 43 images were eventually deemed to be acceptable as test items, although some of these had potentially ambiguous names and might interfere with subjects' retrieval speed. To eliminate such items a pilot study was required before the Pictures test could be used as a reliable measure. That study is described in the next section.

6.5 Study 5: Pictures pilot

6.5.1 Introduction

The purpose of this study was to evaluate the 43 items selected from Powerpoint Clip Art as stimuli in a discrete trial naming speed test. A group of primary age children acted as subjects. It was thought that, if the eventual items were part of primary schoolchildrens' vocabularies, there would be no doubt concerning their suitability for use with secondary age children.

It was considered that items should not be used in the eventual test if subjects were likely to have more than one potential name for the object. It has been demonstrated that objects with "multiple names" such as couch/settee/sofa, take significantly longer to name than objects with high name agreement (Vitkovitch & Tyrrell, 1995; Paivio et al., 1989). It was predicted that certain highly familiar items in the set, such as apple and elephant, would elicit the same name from all subjects. However, some items may have competing alternatives. For example, mug might elicit the name "cup" under experimental conditions.

Another factor under consideration was the ease with which subjects could supply names for the objects. (Initially it had been planned to select age-appropriate items from a standard vocabulary measure.) In order to ensure that the subjects in the eventual studies (Studies 6 and 7) would not find the objects hard to name, it was decided to pilot the stimuli on a younger age group. It is essential for this measure to ensure that the relevant words are in the subject's internal lexicon, since it is speed of retrieval - not vocabulary - which is the variable being measured. Furthermore, it is argued that children's ability to suppress competing names for objects should increase with age (Johnson and Clark, 1988). Therefore items which yield consistent names in a younger sample can be used with confidence with an older sample.

It was also hoped that a number of pictures might be suitable for use as test items that were unlikely to be found in standard vocabulary tests. Powerpoint Clip Art contains a number of technological images which were selected as items for the pilot test. These included an audio cassette tape, a computer terminal, a personal stereo and a satellite dish. Many vocabulary tests still in current educational use feature items (such as spinning tops) which may be unfamiliar to contemporary subjects. If the subjects in this study were able to generate consistent names for the technological items it might have implications for the designers of future vocabulary tests.

A selection of items (including discarded and ambiguous items) is contained in Appendix 3 .

6.5.2 Subjects

The subjects were 25 primary schoolchildren between the ages of 7 and 10 (mean age 8:11).

6.5.3 Method

Each subject was seated in front of the computer and given the same instructions. They were told that they were about to see a series of pictures on the screen and all they had to do was to give the name of the object in each picture. They were told that there was no time limit but that they should say the first name that came into their head. Although this was primarily a test of vocabulary rather than speed of retrieval, subjects were not encouraged to dwell on the task (since this might produce over-elaborate answers). After the subject had named each picture, the experimenter pressed a key which introduced a new picture. Although there was no attempt to measure response latencies, a written record was made of the answers given for each item.

6.5.4 Results & Discussion

As Table 16 shows, a number of items were named consistently by subjects, others less so. 13 items showed a consistency rate of 100%. 6 items showed a consistency rate of 96% (1 inconsistent response). Another 9 achieved fairly high consistency (over 80%). These three sets of items were considered reliable enough to be retained for future use. A number of items were also retained even though their consistency rate was lower. For example, the item "bicycle" was answered consistently by 64% of subjects, and the remainder gave the name "bike". This was considered an acceptable alternative because "bike" is simply an abbreviation, and would not have caused subjects any delay in choosing between competing labels. The same rationale applied to the item "aeroplane" where 28% of subjects responded with "plane". These are consistent with the findings of Vitkovitch & Tyrrell (1995).

The least reliable items were those which elicited a strong alternative, for instance "ship", where 44% of subjects responded "boat", and "microscope" where only 24% of subjects responded with the correct name (the majority responded "telescope"!) The latter finding would seem to justify the choice of using a younger sample in the pilot; if 7-10 year olds cannot identify a microscope, there is no guarantee that 13 year olds will do so with sufficient consistency to risk using it as a test item.

Table 16. Naming consistencies of pilot test items (Study 5)

Target word	Responses	Target word	Responses
Cow	Cow (25)	*Money*	Money (21), others (4)
Audio cassette	Tape (18), others (7)	*Book*	Book (23), others (2)

Microscope	Telescope (11), Microscope (6), others (8)	*Personal stereo*	Walkman (8), Headphones (7), others (10)
Scissors	Scissors (23), Pair of (2)	*Radio*	Radio (17), Tape recorder (4), others (4)
Coin	Coin (19), others (6)	*Umbrella*	Umbrella (25)
Shark	Shark (25)	*Horse*	Horse (25)
Rabbit	Rabbit (24), Hare (1)	*Balloon*	Balloon (25)
Cheese	Cheese (21), others (4)	*Computer*	Computer (25)
House	House (22), others (3)	*Plaster*	Plaster (20), others (5)
Bicycle	Bicycle (16), Bike (9)	*Mug*	Cup (20), Mug (5)
Telephone	Telephone (20), Phone (4), Fax (1)	*Aeroplane*	Aeroplane (16), Plane (7), others (2)
Truck	Lorry (17), others (8)	*Ship*	Ship (14), Boat (11)
Cat	Cat (24), Pussy (1)	*Key*	Key (25)
Apple	Apple (25)	*Fish*	Fish (23), Goldfish (2)
Video cassette	Tape (13), Video (6), others (6)	*Calendar*	Calendar (24), Diary (1)
Football	Football (24), Ball (1)	*Pencil*	Pencil (25)
Helicopter	Helicopter (22), others (3)	*Ambulance*	Ambulance (22), others (3)
Car	Car (23), others (2)	*Clock*	Clock (25)
Tree	Tree (25)	*Watch*	Watch (24), digital (1)
Satellite dish	Satellite dish (8), Satellite (8), others (9)	*Telescope*	Telescope (21), others (4)
Elephant	Elephant (25)	*Train*	Train (25)
Church	Church (24), Chapel (1)		

An interesting finding concerns the higher consistency with which traditional vocabulary test items were responded to. 7-10 year olds are unlikely to have much experience of elephants and sharks in day-to-day living, yet these items produced 100% consistency, while frequently encountered objects such as personal stereos, radios and cassettes elicited a vast range of responses. The only "modern" item which elicited consistent responses was "computer". Altogether 13 items were regarded as too inconsistent, and therefore unreliable, for use in further testing. These were subsequently dropped from the test, leaving a total of 30 items.

Overall, subjects' error rate was remarkably consistent. The average number of naming errors was 10.5, with a standard deviation of 2.0. No subject made more than 14 errors, or less than 8. One or two subjects made a larger number of errors in an attempt to produce precise names (or maybe to impress the examiner). For example, one subject responded "hare" for the item "rabbit" (for no particular reason) and "Ferrari" for "car" (on this occasion, justifiably). One subject was omitted from the analysis since she appeared to have misunderstood the point of the exercise. Instead of responding with object names, she treated the task almost as though it were a word-association task, responding "cutting" for "scissors" and "cute" for "rabbit "!

In general, subjects seemed to have little difficulty following the task instructions, and the experimenter was satisfied with the face validity of the test.

6.6 Summary

In this chapter the development and piloting of two new tests is described. The Windows test is a test of long-term visual recognition memory employing a "yes/no" format in two phases: an inspection phase where subjects view a target set of items, and a test phase where subjects are asked to respond to a mixed set of target and distractor items. A number of items had low discrimination, and these items were removed from the test.

The Pictures test is a test of discrete-trial picture naming, where subjects are required to name a series of pictured objects (e.g. apple, tree). Response latencies are calculated by the use of an internal microphone connected to a voice-activated relay. A study was carried out in which a group of primary school children were asked to generate names for the pictures. On the basis of the consistency of their responses, a number of items were dropped because they elicited too many competing alternative names.

The next chapter describes the application of these tests to a class of 13 year old schoolchildren.

CHAPTER SEVEN: VISUAL RECOGNITION, VERBAL LABELLING AND SPELLING: A CORRELATIONAL STUDY.

7.1 Introduction

The second major study of the project is described in this chapter. Here, the tests constructed and piloted in chapter 6 were administered to a class of 13 year old school children. Also included in the analysis was a measure of Rapid Automatised Naming (RAN) speed. The object of the study was to see if these three measures correlated with spelling ability, as measured by a standard spelling test. However, spelling did not correlate to a significant degree with any measure apart from reading age.

7.2 Introduction to Study 6

Having constructed and piloted the Windows and Pictures tests, it was necessary to administer them to a mixed-ability group of children who would be expected to have reached the age specified for orthographic spelling strategies (Frith, 1985).

In this study, spelling would be assessed by use of a standardised test of spelling ability, the spelling subscale of the British Abilities Scales (Elliott, 1992). This would enable a measure of spelling age to be calculated. Although the Boder word lists were useful for allocating subjects to "good" and "poor" spelling groups, it was thought more appropriate to use a measure which had been standardised on a British population. Furthermore, the spelling subscale of the BAS is based on similar theoretical grounds to the Boder test, in that it contains equal numbers of phonologically regular and irregular words for diagnostic purposes.

It was also decided to test subjects' non-verbal intelligence by means of a standard test. In Study 3, Special Needs teachers were asked to select suitable "poor spellers" whose low ability was, among other things, not attributable to general intellectual deficits. However, when a measure of IQ was obtained for these children, they were found to be mostly below average.

The Windows test had been designed to measure subjects' ability to retain a visual configuration in long-term memory. In Study 4 it was found that verbal labelling did not facilitate performance on the Windows test as it had with the measures of visual sequential memory used in studies 1, 2 and 3. This would appear to support Humphreys and Bruce's (1988) argument that verbal labelling is not a mediating factor in visual recognition.

In chapter 1, it was suggested that words might be retrieved from the internal lexicon as wholes, particularly if their spelling is irregular (Ellis and Young, 1988). Therefore a test of visual recognition ought to show a positive correlation with spelling age as measured by the BAS. However, as with visual sequential memory, it may be that poor visual recognition is only one causal factor of spelling difficulties. Nevertheless there may be individuals for whom poor visual recognition is the primary cause of their spelling difficulties. Such individuals should, therefore, be identified by the Windows test.

The Pictures test is intended to be a separate measure of verbal labelling ability independent of memory. It is devised on the theoretical basis that the speed of retrieval of verbal information is a core deficit in spelling disability. Bear and Barone (1991) argue that rapid retrieval is a necessary skill for attaining the orthographic level of spelling development. To test this hypothesis, two naming speed tasks were included in the test battery for this study. One was a variant on the traditional, manually-administered RAN task. The other (the Pictures test) is a discrete trial naming speed task which measures naming latencies for individual objects. It was possible that both tasks would correlate with spelling age; alternatively, certain individuals might be identified whose spelling difficulties could be attributed to slow speed of lexical retrieval.

7.3.1 Subjects

The subjects were a mixed-ability class of 25 Year 9 children at a comprehensive school in Gloucestershire.

7.3.2 Measures

The following measures were obtained:

- Spelling age (BAS)
- Reading age (BAS)
- Intelligence (Raven's Matrices)
- Visual recognition memory (Windows test)
- Discrete-trial picture naming speed (Pictures test)
- Rapid Automatised Naming (RAN) test

These will now be discussed in some detail.

Spelling age

Having had limited success using the Boder word lists as a spelling measure in Study 3, it was decided to use a standard measure of spelling ability that could be administered easily and include few (if any) unfamiliar words. The spelling subscale of the British Ability

Scales (BAS)(Elliott, 1992) satisfies these criteria. The BAS is a battery of educational and cognitive measures aimed at children up to the age of 15, and is used extensively by educational psychologists as a tool for the assessment of specific learning difficulties. The spelling subscale is also of relevance to the project as a whole, because the BAS manual contains a diagnostic system which is explicitly based on the assumption that specific visual memory deficits relate to corresponding types of spelling error (Elliott, 1992).

In particular, it is claimed that "*children with visual holistic processing difficulties may have consequent difficulties in accurate recall of whole words or even syllables, thus placing an undue reliance on phonetic strategies*" (ibid., p.17). This can be interpreted as saying "Phoenician" spellers ought to perform poorly on visual recognition tasks (viz. the Windows test).

Precise instructions are given for categorising errors as "pre-spelling" (i.e., nowhere near the target word), phonetic and non-phonetic. Quite a wide berth is given for phonetic errors (for instance, "knwo" is considered a phonetically plausible spelling for the word *know*), although not as wide as in some studies (e.g. Goulandris & Snowling, 1991) or the criteria for GFEs in the Boder test (Boder and Jarrico, 1982).

The BAS spelling test comprises 69 items, although a number of short forms of 20 items are outlined in the manual. The raw scores of these tests are converted into "abilities", which can then be converted to a spelling age. It was decided to use Test D, which uses a selection of 20 items scattered throughout the full test. Appendix 4 contains the full list of words in Test D.

Reading age

A measure of reading ability was also obtained by using the appropriate subscale of the BAS. This was obtained because, according to the relevant literature, the RAN test should correlate with reading ability. Failure to replicate this finding might effectively invalidate the association between RAN and spelling.

Clearly it would not be possible to administer a reading test to more than one individual at a time; however, since a distraction activity was required for the Windows test, this created a convenient time slot. The BAS subscale was chosen for its complementarity with the spelling measure. It is a word-reading test rather than a prose-reading test; subjects read out a 5x20 table of unrelated words which increase in difficulty and testing is stopped when all five words in a line are read incorrectly. As with the spelling test, the total raw score is eventually converted to a reading age.

Intelligence

One of the major flaws in Study 3 was the way in which intelligence was assessed. In Study 6 it was decided to assess intelligence using a selected measure administered by the author rather than relying on the school or LEA to provide data. An advantage of using a correlational design is that there is no need to screen the sample and IQ can simply be entered as another variable.

Given that general measures of IQ contain a verbal IQ component, it was felt that it might be more instructive on this occasion to obtain a measure of *non*-verbal intelligence. This was based on the idea that children with literacy deficits would be unlikely to excel on a measure of verbal IQ and that a standard measure of abstract reasoning would be more likely to provide a true reflection of intellectual ability.

Raven's Progressive Matrices (Raven et al., 1983) "*seem to be the world's longest and most widely established tests of intelligence*" (Turner, 1994, p.20). They are non-verbal in that instructions are given orally and subjects are simply required to indicate their answer by writing the appropriate number on a form. The task itself involves studying a sequence of patterns in which one box is left blank and having to select the correct missing pattern from several alternatives. The patterns are arranged so that the correct missing pattern can be identified via a series of logical deductions.

There are several variants on the basic theme: for children, the appropriate test is the Standard Progressive Matrices. This comprises 60 items and takes an hour to administer. However, Turner (1994) has recently streamlined the test, creating a short form of only 20 items. It was decided to administer these to the whole class in the same session as the spelling test. Three practice items were selected from the full range, for use prior to the test itself to familiarise subjects with the task instructions. The raw scores were multiplied by three to allow comparison with the age norms provided in the manual.

Visual recognition

The Windows test was modified as described in section 6.3.4 of the previous chapter. The test had been reduced to 28 items, and was administered to individual subjects using the same procedure as in Study 4 (outlined in section 6.3.3).

Discrete-trial picture naming speed

The Pictures test, having been reduced to 30 items, was administered using the same procedure as in Study 5 (see section 6.5.3). Prior to the test itself, each subject received five practice items which would enable them to understand the task requirements.

Rapid Automatised Naming (RAN)

A RAN task was devised using the basic format employed by Denckla & Rudel (1976). Five different objects were repeated 10 times at random in a 5 x 10 matrix. The objects were simple line drawings of five objects - mouse, boat, drum, candle, ladder (see Appendix 5 for reproductions). Although one or two of these objects might have elicited alternative labels (e.g., "rat"), it was considered that any hesitation due to competing labels would account for only an insignificant portion of the total time. Hesitation would only occur once, while naming the first row of objects, and so to nullify this effect, the stopwatch was only started when the first row had been named. Thus the first row was not included in the overall time. A practice row of objects (different from the test items) was given to subjects prior to the full RAN test. This was done to ensure that subjects were aware of the task requirements.

It was expected that RAN times should correlate negatively with reading and spelling ages. It was also expected that RAN times should correlate positively with the Pictures test, which would suggest that the two tasks tap the same skill.

7.3.3 Procedure

Each child was excused from class for a period of between 30 and 45 minutes. They were taken to a small office in the school where testing was able to take place without any interruptions or noise disturbance. The order of testing was as follows:

- Pictures test
- Windows inspection phase
- BAS reading test
- Windows test
- RAN test

The spelling test and Raven's Matrices were administered separately to each subject.

7.4 Results & Discussion

Table 17 contains the means for each measure, along with standard deviations and minimum and maximum values. The full correlation matrix for the six variables is presented in Table 18. Significance levels were arrived at using the tables provided by Powell (1976). The results will now be discussed in three sections: spelling/reading/IQ (general educational variables); Windows results (visual recognition); Pictures and RAN results (picture naming speed).

Table 17. Means, SDs, and maximum/minimum values (Study 6)

Variable	Means	SDs	Range (min - max)
Chron. age	13:3	0.4	12:9 - 14
Spelling age	11:1	1.4	8 - 13:6
Reading age	11:6	1.7	9 - 14:5
Raven's (adjusted)	39.5 (/60)	6.2	24 - 51
Windows	16.9 (/28)	3.3	11 -23
Pictures	1.32 (secs)	0.3	1.04 - 2.68
RAN	29.7 (secs)	6.7	18.4 - 51.8

Table 18. Correlation matrix (Pearson's *r*) for Study 6

	Spelling	Reading	Raven's	Wins	Pictures	RAN
Spelling	1.00					
Reading	0.594**	1.00				
Raven's	0.207	0.264	1.00			
Wins	-0.237	-0.424*	0.179	1.00		
Pictures	0.007	-0.143	0.457*	0.024	1.00	
RAN	-0.038	-0.104	-0.348	-0.095	0.016	1.00

** Significant at the 0.01 level

* Significant at the 0.05 level

7.4.1 Analysis of individual measures

Spelling, Reading and IQ

It can be seen from Table 18 that the only variable to correlate significantly with spelling was reading, which is an expected finding. However, an unexpected finding was that the group as a whole were found to have below average ability in both reading and spelling (according to the norms established for the BAS). Indeed, subjects' spelling ability was on average 2 years behind what would be expected. Only one subject was found to have a spelling age above her chronological age. The implication of these findings is that this group is not truly representative of the general population. However, the range of spelling ages is wide (almost 6 years), which means that it is still possible to draw some conclusions from the data.

The Raven's scores also suggest that this group performs as a whole below the norms established for its age-range. However, this masks the fact that there is considerable variation within the sample, including one subject whose score fell above the 95th percentile for her age group. Unlike in Study 3 there is no correlation between spelling ability and intelligence, though it is difficult to say whether this is due to the design of the study, or to the use of a non-verbal intelligence test.

From an administration point of view, the BAS subscales were a successful means of establishing spelling age. Unlike the Boder word lists, these test contained no problematic items. It was thought that Test D was not the ideal choice for a short form for this age group, since the first 10 items were generally too easy; in a below-average ability sample, very few subjects mis-spelled these items. This meant that subjects' raw scores were contained within a narrow band, and yet their conversion to spelling ages made it appear a wide range. An associated problem is the lack of error data from which to categorise subjects into "Phoenician" and "Chinese" groups. Nevertheless, an error analysis of this sort was undertaken: 20 subjects were allocated to the Phoenician group and 5 to the Chinese group.

Windows test

The revised Windows test appeared to be more difficult than the piloted version. The average "hit rate" for this group was 60 per cent, compared to 70 per cent in the pilot. This trend is the reverse of what would be expected (given the reduced memory load). However, the difference in age between the two groups must be borne in mind (and the fact that the easiest items had been removed).

It was expected that, given the prediction based on Elliott's (1992) diagnostic system, "Chinese" spellers would perform better than "Phoenician" spellers on the Windows. The rationale for this prediction was that the holistic visual processing of Phoenician spellers was deficient and forced them to rely too heavily on phonetic strategies. Chinese spellers did in fact perform better (a mean of 17.8 compared with 16.7), but this result did not attain significance ($t = 1.8$, d.f. $= 23$, $p > 0.05$). However, the small number of Chinese spellers makes it difficult to glean any information from means comparisons.

Picture naming tests

The Pictures and RAN tests did not correlate with each other, reflecting the findings of Stanovich *et al.* (1983), though not Wolff *et al.* (1990). This might lead one to conclude that they are not testing a common ability. Furthermore, the only significant correlation involving either measure was a positive correlation between the Pictures and Raven's scores. This suggests that the more intelligent subjects actually took *longer* to name the pictures.

Such a finding might prompt the argument that the more intelligent subjects were deliberating longer because they were seeking more precise definitions. However, inspection of the Pictures data reveals that only 2 of all 750 items elicited names that were different from those expected by the experimenter. If intelligent subjects were searching for precise definitions they were not finding them. Conversely, the RAN times were correlated negatively with IQ (although not significantly). The disparity between these findings lends support to the argument that RAN is a measure of articulacy rather than processing speed.

Table 19 displays the mean response latency for each item in the Pictures test. Although there appears to be a slight effect favouring traditional test items (apple, cat, etc.) the range of times is fairly narrow, suggesting that subjects did not have difficulty naming any item (all means are inside 2 seconds). There still appears to be a slight bias in favour of traditional vocabulary items. For example, apple (1.06) and cat (1.13) were named faster than telephone (1.7) and calendar (1.53). Given that there were only two errors made in the entire sample, there seems little doubt that the final selection of items are unambiguous representations of the depicted objects.

The Pictures failed to correlate at all with spelling (in fact the correlation is almost zero). This suggests that, when isolated from its role as a strategy in memory tests verbal labelling does not appear to be a successful predictor of spelling ability. However, before that conclusion can be drawn it is necessary to consider other possibilities.

Table 19. Mean response times (secs) for Pictures test items

Cow	1.49	*Book*	1.12
Bicycle	1.16	*Computer*	1.39
Apple	1.06	*Horse*	1.16
Balloon	1.13	*Scissors*	1.21
Shark	1.54	*Telephone*	1.70
Rabbit	1.21	*Watch*	1.53
Umbrella	1.38	*Calendar*	1.53
Church	1.32	*Fish*	1.52
Tree	1.22	*Key*	1.20
Car	1.38	*Pencil*	1.14
Football	1.18	*Clock*	1.25
Elephant	1.16	*Train*	1.24
Cat	1.13		

The first concerns the role of the RAN task in this study. Given that the task was carried out in accordance with the procedures established in the literature, the lack of a strong relationship with reading can perhaps be attributed to the age of the subjects. Most of the studies using RAN tasks have been carried out using a younger population (primary age children); only Wolff et al. (1990) claim to find a significant result using older children. Furthermore, the majority of studies using RAN are controlled experiments comparing normal spellers with "dyslexic" subjects. No studies known to the author have looked at correlations within a normal population.

Another possibility is that correlational data may hide some individual cases where slow speed of lexical retrieval is a factor affecting spelling development. Wilson and Cline (1995) found that, although naming speed appeared to be a good predictor of reading ability, there were individual cases of poor readers with very fast naming speed. The authors argue that, although slow naming speed is often associated with delayed reading development, the relationship is not a linear one. Therefore, as with the visual sequential memory measures used in Study 3, naming speed tasks may only account for spelling difficulties in a number of cases. This possibility is considered in more detail in the next section.

7.4.2 Single case studies

As with Study 3, a number of individual cases were selected for analysis. Table 20 displays their chronological, reading and spelling ages and performance on the Windows, Pictures and RAN tests.

Table 20. Single case data for Study 6. The italicised numbers in brackets represent individual deviations from the overall mean. The italicised letters in brackets indicate whether the subject's reading and spelling errors were classified as "Phoenician" or "Chinese".

Subject	CA	RA	SA	IQ	Windows	Pics (s)	RAN (s)
MD	13:2	14:5 (*P*)	10:6 (*P*)	33 (*-6*)	11 (*-5*)	1.27 (*+.03*)	31.5 (*+2.7*)
GY	13:1	9:9 (*P*)	10:6 (*C*)	45 (*+6*)	14 (*-2*)	2.68 (*+1.44*)	40.5 (*+10.7*)

MD, whose Windows score was the lowest in the sample, appears to belong to the "good readers-poor spellers" subgroup identified by Frith (1980). Most of his spelling errors were good phonetic equivalents, for example "sitt" for *sit* and "begining" for *beginning*. This pattern is also consistent with the theory of Elliott (1992), that subjects with poor "visual processing" skills would display an *undue reliance on phonetic strategies*" (ibid., p.17). However MD's reading performance was above average for his age group. Frith's argument is that it may be possible to read using "partial cues", where enough letters are identified to enable the reader to make a guess based on contextual information. However spelling requires more precise recognition, a skill which may have been tapped by the Windows test.

GY's score on the Windows test is also below the sample mean. However his scores on the naming speed tests are perhaps more indicative of his poor reading and spelling difficulties. His spelling errors are "Chinese", such as "beiging" for beginning, suggesting a difficulty with phonological processing. It appears that this difficulty may have been identified by the slow naming speed times.

A third case study, CR, showed no particular pattern of performance, although her reading and spelling were well below the mean expected by her age. Her Windows score was above the sample mean, which might suggest that, in her case, poor visual recognition is not an identifiable cause of spelling difficulty. Such a suggestion may be borne out by CR's spelling answer sheet, on which several mis-spellings are crossed out and replaced with correct spellings. In the case of this individual, the Windows test seems to have identified competence in visual recognition.

7.5 Conclusion and general remarks

In this chapter, the Windows and Pictures tests were administered to a mixed-ability group of 13 year olds. Neither test correlated with spelling age as measured by the spelling subtest of the British Ability Scales. This may suggest that visual recognition, and speed of lexical access, are not factors which are related to spelling in themselves, although the spelling difficulties of certain individuals can perhaps be traced to deficits in either of these processes, as case studies appear to suggest.

The use of the BAS reading and spelling tests appeared to be a successful way of estimating the spelling age of the subjects in the study, although Test D, the short form, may have contained too many easy items for 13 year old subjects. Therefore, subjects were discriminated on the basis of only one or two items. In the next study it was decided to use Test C, which comprises the 20 most difficult items from the full test.

The Windows test failed to correlate with spelling, although in two individuals it appeared to be related to spelling performance. As with the Triangles test, it is argued that the skill tapped may not be essential for spelling competence by itself; however, a notable deficit in visual recognition may manifest itself in poor spelling in certain individuals.

The same argument can be applied to the Pictures test, with respect to speed of lexical access. Certain individuals may display a deficit in retrieving object names quickly, and this may be related to poor phonological processing, which in turn may manifest itself in spelling difficulties. In general, however, speed of lexical access does not appear to be related to spelling ability.

One major purpose of Study 6 was to separate the component skills of visual memory and verbal labelling, which together - in the Animals test - discriminated between good and poor spellers in Study 3. The general conclusion to be drawn from Study 6, therefore, is that neither of these component skills by themselves are related to spelling except in a few individuals.

CHAPTER EIGHT: VISUAL MEMORY, VERBAL LABELLING AND SPELLING: A SPELLING-LEVEL DESIGN STUDY.

8.1 Introduction

The third and final major study of the project is described in this chapter. In Study 6, visual memory and verbal labelling were tested separately, by use of the Windows and Pictures tests. In Study 7 it was decided to use these as independent variables in a spelling-level design study, along with the Triangles and Animals tests used in Study 3. This would ensure that both aspects of visual memory - visual sequential memory and visual recognition - would be tested using the same sample, along with a measure of verbal labelling (the Pictures) and a combination of the two (the Animals). The advantages and disadvantages of using such a design are explained in some detail. Subjects were screened before selection so that the three groups had mean IQ close to 100, and that the experimental group had a mean spelling age at least two years behind mean chronological age. Experimental subjects differed from controls on only one measure, the Animals test, in which poor spellers performed less well, thus replicating the finding from Study 3. It is argued that, on the basis of these results, that visual memory is only related to spelling ability for certain subjects. The integration of visual and verbal information - which could be explained in terms of automaticity - would appear to be a more important process with respect to spelling.

8.2 Introduction to Study 7

The tests designed for this project have been used to investigate the following cognitive processes:

Animals test Visual sequential memory, mediated by verbal labelling

Triangles test Visual sequential memory, independent of verbal labelling

Windows test Visual recognition, independent of verbal labelling

Pictures test Verbal labelling, independent of visual memory

The general indication in Studies 3 and 6 was that the Animals test is the only test which favours normal spellers over poor spellers. Individual poor spellers may demonstrate weaknesses in the skills tapped by the other tests, but the skills tapped by the Animals test appear to be a basic requirement for competent spelling. It was suggested that these skills may be defined as "verbal labelling", and the Pictures test attempted to isolate these skills from a memory test context. Verbal labelling, or speed of lexical access, failed to correlate with spelling age when isolated in this way. Similarly, tests of visual memory which do not appear to be mediated by verbal labelling - the Triangles and Windows tests - have not been shown to favour good spellers over poor ones. Therefore it may seem that the pattern of results produced by the Animals test is related to normal spellers' ability to combine verbal and visual information.

In order to test this hypothesis fully, it was necessary to use all four measures in a controlled study comparing poor spellers with controls matched for spelling age and chronological age.

Study design

It was decided to administer the above four tests in a spelling-level design study in which it would be possible to control for some of the measures which had hitherto been problematic, such as IQ. This was based on the reading-level design advocated by Backman, Mamen & Ferguson (1984), in which poor readers are compared with two sets of controls, one group matched on chronological age and the other matched on reading age. It was also intended to use all the tests devised for the project, which would enable the tests to be compared with one another and to be used as independent measures comparing poor spellers with controls matched for both age and spelling ability.

Backman et al. (1984) advocate a reading-level design because, they argue, positive results in a traditional chronological age (CA) match study do not reveal causal factors. With reference to the current project, doubts must exist concerning the superiority of CA controls on tests such as the Animals, which has led to the suggestion that verbal labelling is an important requirement for spelling. Yet this result might indicate, as argued in Chapter 5, that verbal labelling is simply a "by-product of literary attainment" and not a causal factor in its own right. In order to control for achievement-related factors it is necessary to obtain a measure based on a group matched for ability.

However, the reading-level (RL) design has been criticised by Bryant & Goswami (1986), who argue that negative results obtained by this type of design may mask genuine causal factors. For example, if RL matched groups failed to differ on a memory task, this may be due to the fact that the poor readers have developed strategies to cope with memory difficulties. Another possibility is that cognitive task performance improves with age because older children have more experience of performing under test conditions, and that a practice effect emerges (Anastasi, 1990).

Therefore it was decided that, if a spelling level design were to be used for Study 7, intergroup differences should only be regarded when the direction of results is consistent with both control groups. In other words, if a measure proves positive with one group and not the other, it should not be regarded as a valid finding.

Spelling measure

As in Study 6, it was decided to use the spelling and reading subscales of the British Ability Scales as a measure of spelling and reading ability. In that study, Test D was used as the spelling measure, a short form created by selecting every third item from the full 60-item test. However, since the BAS is only designed for use with children up to the age of 15, this meant that most of the items were too easy for a 13 year old sample. Test C, which comprises the most difficult 20 items, was felt to be a more suitable measure of spelling ability for this age group, and so this short form was the one adopted.

In Chapter 7, it was argued that the BAS was a suitable test for Study 6 because the theoretical basis of its diagnostic system is relevant to the present project. Children with poor "visual processing", it was argued, will rely heavily on phonetic strategies (Elliott, 1992). This was supported by one case study, MD, but not by another, GY. Correspondingly, "*Children with sequential processing difficulties may have consequent difficulties in using phonetic strategies in spelling*" (ibid., p.17). This can be interpreted (within the scope of this project) as saying: "Chinese" spellers ought to perform poorly on visual sequential memory tests (e.g. the Animals and Triangles tests).

Hypotheses

The following predictions were made, based on the findings of the previous studies.

- The Animals test would discriminate in favour of both control groups, replicating the findings of Study 3. This is because verbal labelling is a facilitatory strategy and poor spellers appear to be disadvantaged on memory tasks which involve verbal labelling. This is consistent with the findings of Hicks (1980), H.L. Swanson (1984), L. Swanson (1978) and others who have demonstrated that poor *readers* are less adept at using a verbal labelling strategy unless prompted. However, a measure of verbal labelling as a separate process from visual memory did not correlate with spelling age in Study 6. This suggests that it may be the *combination* of visual memory and verbal labelling which poor spellers find difficult.

- The Triangles test would produce no significant intergroup differences. On the basis of Study 3, it appears that visual sequential memory is not a factor which discriminates between good and poor spellers. This runs counter to the arguments of Link and Caramazza (1994), Shallice (1988), and Thomson (1984) and the empirical findings of Bryant & Bradley (1981). The last of those studies used verbal stimuli. By using abstract stimuli which has been found to be difficult to label verbally, this test is not expected to favour controls

- The Windows test would produce no significant intergroup differences. As with the Triangles test, verbal labelling does not appear to facilitate performance on this task. Humphreys & Bruce (1989) argue that this is because labelling is less advantageous in tests of visual recognition. Therefore, as in Study 6, it is unlikely to discriminate between good and poor spellers. This would run counter to the findings of Ormrod (1985) and Tenney (1980) who argued that poor spellers have poor recognition skills. Again, however, the stimuli used in those studies was of a verbal nature.

- The Pictures test would not produce significant intergroup differences, since verbal labelling does not appear to discriminate between good and poor spellers except as a sub-skill on a memory test (e.g. the Animals test). The results of the Pictures test in Study 6 did not appear to support the suggestion by Bear and Barone (1991) that rapid retrieval was a necessary skill for reaching the orthographic spelling stage, since no correlation was found with spelling age. However, as Wilson and Cline (1995) argue, correlational data may hide subjects for whom slow access of lexical retrieval is a major factor in spelling difficulty. Investigation of single cases in Study 6 appears to lend some support to this argument.

8.3.1 Subjects

Three groups, each comprising 20 subjects, took part in the study. These were selected to conform to the following criteria:

Group 1 (Experimental):	Mean age 13, Spelling age 11, IQ 100.
Group 2 (CA control):	Mean age 13, Spelling age 13, IQ 100.
Group 3 (SA control):	Mean age 11, Spelling age 11, IQ 100.

Pupils in years 7 to 9, from 15 secondary schools covering a wide geographical area of the West Midlands formed the sampling frame for this study. The LEA agreed to co-operate and provide ability test scores for those pupils who had been given the Cognitive Abilities Test on entry to secondary school. This list was used to screen out pupils of low ability, and 7 schools were used to form the sample. Examination of the IQ scores for these schools indicated they were representative of the 15 in the sampling frame.

All pupils in Year 7 to Year 9 were then given the BAS spelling test, and the eventual groups were selected from the BAS scores and IQ scores. The means for the three groups are displayed in Table 21.

Table 21. Mean age, spelling and IQ data for the three groups (Study 7)

Groups (n = 20)	CA	SD	SA	SD	IQ	SD
1 (Expt)	**13:4**	0.5	**10:6**	0.7	**99.1**	5.6
2 (CA match)	**13:3**	0.4	**13:2**	0.9	**102.8**	9.5
3 (SL match)	**11:2**	0.3	**11:1**	1.0	**101.5**	4.5

From the above set of figures it can be seen that the three groups correspond to the required criteria, although the spelling age of the experimental group is slightly below that of the SA controls. IQ did not differ significantly among the three groups ($F_{1,59} = 1.56$, $p > 0.05$), so it was not deemed necessary to use a covariant for the statistical analyses.

8.3.2 Measures

The following measures were obtained:

- Reading age (BAS)

- Visual sequential memory (Animals & Triangles tests). The same set of tests as used in Study 3, although omitting the Kirk test. The rationale for this is that the Kirk test was created initially to examine the facilitative effect of verbal labelling, and to act as a template for the other two tests by using standardised stimuli. It was deemed to have served these two purposes, although it was felt that its discriminative powers were not strong enough for any further use.

- Visual recognition memory (Windows test). The 28-item test, as used in Study 6.

- Discrete-trial picture naming speed (Pictures test). As in Study 6.

8.3.3 Procedure

The tests were administered on an individual basis. Each child was excused from class for a 20 minute period. As in Studies 3 and 6, subjects were taken to a small office in the school where testing was able to take place without any interruptions or noise disturbance. The order of testing was as follows:

- Pictures test
- Windows inspection phase
- Windows test
- Animals & Triangles tests

8.4 Results and Discussion

Table 22 displays the means and standard deviations for the four measures in the study. Performance across the three groups on each measure was compared using analysis of variance. No significant differences between the groups could be found on any measure other than the Animals test, a finding which confirmed the prediction outlined in section 8.2.

Table 22. Intergroup comparisons (Study 7)

Test	Group 1 (n = 20) Experimental		Group 2 (n = 20) CA controls		Group 3 (n = 20) SA controls	
	M	SD	M	SD	M	SD
Animals (/15)	**9.8**	2.1	**11.6****	2.1	**11.8****	2.5
Triangles (/15)	**9.6**	2.6	**10.3**	1.7	**9.9**	2.1
Windows (/28)	**18.5**	3.0	**18.4**	2.3	**17.7**	4.1
Pictures (ms)	**698**	213	**741**	205	**821**	218

** Significant at the .05 level

The Animals test produced a significant difference between the groups ($F_{2,47} = 5.33$, $p < 0.01$). Scheffé's multiple comparison test found a significant difference in means at the 0.05 level for the poor spellers with both control groups. This suggests that the intergroup difference in performance on this test could not be attributed to maturational factors. Therefore the doubts expressed by Bryant and Goswami (1986) about the reading-level design were not applicable to this study.

The Triangles test produced no significant differences between the groups ($F_{2,47} = 0.46$, $p > 0.05$), nor did the Windows test ($F_{2,47} = 0.3$, $p > 0.05$). These results confirm the prediction that a measure of visual memory would only differentiate between the groups when verbal labelling could facilitate task performance.

As predicted, the Pictures test - a measure of verbal labelling independent of visual memory also failed to produce a significant difference ($F_{2,47} = 1.36$, $p > 0.05$). Therefore poor spellers' poor performance on the Animals task cannot be attributed to the fact that slow rate of lexical access makes verbal labelling ineffective.

This pattern of results can be interpreted from three perspectives. Firstly, it could be argued that the poor spellers, like the dyslexic subjects in the Hicks (1980) study, are failing to capitalise on a verbal labelling strategy in the Animals test because it is not their preferred mode of operation. This could be described, following Bjorklund and Coyle (1995), as strategy utilisation deficiency. Such an explanation is consistent with the argument of Lennox and Siegel (1994), that poor spellers have difficulty integrating visual and phonological information.

The second possibility is that this failure is due to an inability to function within two modalities simultaneously. When visual memory and verbal labelling are separated, as in the other measures in the study, poor spellers perform at the level of the controls. It is the *combination* of these skills that produces a decrement in performance. Such an explanation could be interpreted in terms of the dyslexic automatisation deficit theory of dyslexia (Nicolson & Fawcett, 1990,1995). This possibility is discussed in more detail in section 9.5.

A third perspective from which to view the poor spellers' difficulty in the Animals test is in terms of computational resources. Brown and Loosemore (1994) argue that lack of computational resources is a core deficit in developmental dyslexia. In terms of Baddeley and Hitch's (1974) working memory model, it could be argued that the deficit lies at the level of the central executive, the "processing space" where visual and phonological information is integrated (via the visuo-spatial scratchpad and the phonological loop).

The failure of the Triangles and Windows tests to differentiate between poor spellers and controls has some implications for cognitive theories of spelling. Both tests have been carefully designed to tap some of the processing skills that, it has been claimed, are involved in spelling (Thomson, 1984; Shallice, 1988; Elliott, 1992; Link and Caramazza, 1994). However, unlike previous measures administered in that field (e.g., Ormrod, 1985; Bradley & Bryant, 1981), they employ stimuli that are non-verbal in nature. In addition, neither test appears to be facilitated by use of a verbal labelling strategy. On the basis of these results, it appears that the cognitive processes in spelling are more complex than has been assumed.

Poor Spellers subgroups

As in the previous studies, the poor spellers were divided into two groups on the basis of spelling error. Again, two independent referees validated the author's classification ($r = 0.71$ and 0.68 respectively, both $p < 0.01$). These subgroups did not differ on IQ or spelling ability ($t = 0.2$ and 0.1 respectively). Table 23 displays their performance on the independent measures.

Table 23: Comparison of Phoenician and Chinese spellers (Study 7)

Test	**Phoenicians (n = 12)**		**Chinese (n = 8)**	
	M	SD	M	SD
Animals	**10.2**	2.2	**10.0**	2.2
Triangles	**9.8**	2.6	**9.4**	2.7
Windows (scores)	**19.4**	3.0	**17.3**	2.5
Pictures (ms)	**655**	200	**763**	228

Phoenician spellers might have been expected to display an advantage on tests involving phonological material, but although their mean time in the Pictures test was faster than that of the Chinese spellers, the difference was not significant ($F_{(1,19)} = 1.2$, $p > 0.05$). Neither did the Animals test discriminate between the two subgroups ($F_{(1,19)} = 0.1$, $p > 0.05$).

Chinese spellers might have been expected to display an advantage on tests where sequential processing was involved (Elliott, 1992), but the Triangles test did not discriminate between the subgroups ($F_{(1,19)} = 0.1$, $p > 0.05$). Finally, the higher mean score of the Phoenicians on the Windows test was not found to be significant ($F_{(1,19)} = 2.7$, $p > 0.05$).

This set of results suggests that the pattern of errors produced in a spelling test may be the result of a vast range of cognitive factors, and that the subtype theories of Boder (1973) and Treiman (1984) may be an over-simplification. In the next paragraph, further evidence for this argument is presented in the form of detailed case studies.

Case Studies

Table 24 contains the reading and spelling ages and IQ for three selected subjects from the Poor Spellers group, along with their scores on the Animals, Triangles and Windows tests.

Table 24. Single case data for Study 7. The italicised numbers in brackets represent individual deviations from the overall mean. The italicised letters in brackets indicate whether the subject's reading and spelling errors were classified as "Phoenician" or "Chinese".

Subject	CA	RA	SA	IQ	Animals	Triangles	Windows
JS	13:2	10:6 (*P*)	10:4 (*P*)	98	10 (-*1*)	5 (-*5*)	12 (-*6*)
GB	13:1	10:10 (*P*)	11 (*P*)	94	12 (+*1*)	6 (-*4*)	21 (+*3*)
DA	12:7	11:7 (*P*)	10:5 (*C*)	110	6 (-*5*)	6 (-*4*)	16 (-*2*)

Subjects JS and GB make an interesting comparison. Both are reading and spelling at more than two years below their chronological age despite near-normal IQ. Both performed on the Animals test at a level close to the whole sample mean, but were well below the sample mean on the Triangles test. However, JS performed well below the sample mean on the Windows test while GB performed slightly above the sample mean. It could be argued that, on the basis of these results, JS has poor visual recognition *and* visual sequential memory but GB has only poor visual sequential memory. Analysis of their spelling errors shows that GB makes several crossings out in an attempt to arrive at the correct spelling for some words. Like CR in Study 6, this appears to indicate an ability to recognise mis-spellings. JS seems to lack this ability; this may account for her lower spelling age.

These case studies arguably demonstrate the way in which good spellers may combine visual recognition and visual sequential memory. Deficits in one area (e.g. visual sequential memory) may be sufficient to lead to spelling difficulties but can be alleviated by compensatory strategies, such as correcting mis-spellings once written down. Deficits in *both* areas may mean that the subject takes longer to reach the "orthographic" stage of spelling development (Frith, 1985).

DA is the subject in the sample who comes closest to Frith's "good readers-poor spellers" subgroup. Unlike MD in Study 6, his spelling errors follow the "Chinese" pattern, a typical example being his spelling of his date of birth ("Arpil 22"). This suggests that, like HW in Study 3, his principal difficulty lies at the level of visual sequential memory. This is indicated by his poor performance on both Animals and Triangles tests. The low Animals score may suggest that his phonological processing abilities are not sufficiently developed to enable him to attempt good phonetic equivalent spellings.

As in the previous studies, analysis of individual cases highlights the way in which the measures devised may identify deficits which account for some spelling difficulties. They also demonstrate the complexity of spelling disability and the problems involved with attempting to identify deficits which are common to all poor spellers.

8.5 Summary and conclusions

Study 7 was an attempt to obtain measures on all the tests devised in this project and investigate comparisons between poor spellers and controls on these tests. A spelling-level design was favoured, so that it would be possible to rule out maturational factors and the side-effects of literacy attainment. All three groups were matched for intelligence. The only test that discriminated successfully between the three groups was the Animals test, in which both CA and SA controls performed significantly better than the experimental group. This replicated the findings of Study 3, and was interpreted as evidence for poor spellers' lack of effective strategy use, possibly due to an inability to "automatise" separate sub-skills. As in Studies 3 and 6, analysis of single cases appeared to identify specific individuals whose spelling difficulties could be accounted for by weaknesses in the skills tapped by the various tests.

CHAPTER NINE: GENERAL DISCUSSION

9.1 Introduction

This chapter begins with a brief summary of the project findings. In section 9.3 these are discussed with reference to the dual route model of spelling discussed in Chapter 1, and to subtype theories of dyslexia. Section 9.4 provides an evaluation of the original measures used in the project - the Animals, Triangles, Windows and Pictures tests. Only one of these measures, the Animals, differentiated between good and poor spellers, and the theoretical implications of this finding are discussed in section 9.5. The poor performance of poor spellers on this test is explained in terms of automaticity theory and in terms of effective strategy use. It is argued that a supervisory attentional system may be implicated in both these processes. Finally, in section 9.6, a number of suggestions are made regarding possible directions for future research.

9.2 Summary of the project findings

The basic aim of the project was to examine the role played by visual memory in the spelling of 13 year olds. This age group was focused on because it has been suggested that orthographic information is more important at this stage of spelling development (Frith, 1985). It was decided to devise a number of original measures which would tap the visual memory skills that might be involved in the spelling process (visual sequential memory and visual recognition).

One major question addressed was the role of verbal labelling as a strategy that facilitates performance on visual memory tests. It was found (in studies 1 and 2) that, in a normal subject population, easily-labelled stimuli were easier to recall than stimuli that were difficult to label. When visual sequential memory tests were administered to good and poor spellers (Study 3), the groups showed no difference in performance using stimuli that were hard to label (the Triangles test). However, good spellers achieved significantly higher levels of performance on a test using easily-labelled stimuli (the Animals test).

It was suggested therefore that poor spellers might experience difficulties with verbal labelling itself; this had been described as the core deficit in dyslexia (Miles, 1993). In Studies 6 and 7, a test of verbal labelling was created that did not involve a memory element, recording the time subjects took to supply a verbal label for a familiar picture (the Pictures test). This measure failed to differentiate poor spellers from good ones.

In the final study of the project (Study 7) poor spellers were compared with chronological-age and spelling-age matched controls on measures of visual sequential memory, visual recognition and naming speed (verbal labelling). It was found that the only measure that favoured both control groups over the poor spellers was the Animals test. It was suggested that this measure differed from the others in the project because it required subjects to operate in both visual and verbal modalities. It was also suggested that the various measures may be of some use in identifying possible causes of spelling retardation in individual subjects.

9.3 Implications for theories of spelling

In the next section, the implications of the above findings are discussed with reference to two theories of spelling which have been reviewed earlier in the thesis. To begin with, the implications are discussed for the dual route spelling model. Secondly, there is a discussion of subtype theories of spelling disability and the use of error analysis as a means of identifying cognitive processes.

9.3.1 The dual route model of spelling

In most models advocating a dual route of spelling (Ellis and Young, 1988; Shallice, 1988; Link and Caramazza, 1994; Barry, 1994) it is postulated that normal spellers possess four "internal lexicons" of known spellings, one of which is a "graphemic output lexicon". This hypothetical structure enables an individual to spell a word without recourse to phonological information. It is argued (Ellis and Young, 1988) that this is achieved by retrieving the word as a whole from memory rather than piecing it together using the phonological rules of language. Studies of brain-damaged individuals suggest that this is how words are spelled when the phonological spelling route is disrupted (Shallice, 1981; Bub and Kertesz, 1992).

It follows therefore that an individual with a poor visual memory will have to rely heavily on sound-spelling rules, a spelling route which may be of little use in spelling homophones or phonologically irregular words such as *yacht*. In studies 3, 7 and 8 a number of individuals were identified with low scores on visual memory tests who appeared to rely too heavily on phonological information in their spellings. Further subjects were found to have deficits in visual memory *and* phonological processing (as measured by the Pictures and Animals tests), which caused them to make errors which were nonphonetic.

It was suggested that two basic visual memory processes are required for spelling: visual sequential memory and visual recognition. Visual sequential memory is used to recall the precise letter order of a word, and has been modelled with limited success by Houghton, Glasspool and Shallice (1994). Visual recognition can be used to check the spelling of a word once it has been written; indeed two subjects (CR in Study 7 and GB in Study 8) scored highly on the Windows test, and used their apparently good visual recognition in order to cross out mis-spellings and substitute correct versions.

The findings of the project have no direct bearing on the dual route model of spelling; however they suggest possible ways in which information might be retrieved from the graphemic input lexicon, and possible constraints on the use of this information. Shallice (1988) suggests that, following its retrieval from the lexicon, a word is held in a "*visual short-term store*" (p.156) which could be defined as working memory (Baddeley and Hitch, 1974). At this point in the spelling process, visual sequential memory might be used to recall the correct letter order. However, the initial retrieval relies on some form of visual imagery, a process which has not been investigated in this project.

Although the identification of a number of single cases from the studies points to selective deficits in various aspects of visual memory and its role in their spelling, it was not possible to find a *general* relationship between visual memory and spelling in the overall samples. It can be argued, therefore, that in 13 year olds at least, visual memory plays a complex and elusive role in spelling that may only be identified at the level of the individual.

9.3.2 Subtype theories of dyslexia

In Chapter 1 a number of "subtype" theories of dyslexia were reviewed. These theories argue that there are two or more distinct varieties of spelling disability, which can be identified by the analysis of spelling errors found using specially constructed spelling tests. One such example is Boder's (1973) theory in which dyslexic subjects are classified as "dysphonetic" (phonological processing deficit), "dyseidetic" (visual memory deficit) or "mixed" (processing deficits at both levels). Treiman (1984) and Weekes (1994) have put forward similar proposals; Treiman (1984) argues that dyslexic spellers can be ranked along a continuum ranging from "Chinese" spellers (phonological processing deficit) to "Phoenician" spellers (visual memory deficit). Other researchers have identified subgroups of 13 year olds whose reading is at a near-normal level but whose spelling age lags over two years behind their chronological age (Frith, 1980; Batchelor *et al.*, 1990; Newman *et al.*, 1993).

The case studies reported in Studies 3, 7 and 8 seem to identify individuals who fall into all the subgroups mentioned in the above paragraph. In Study 7, for example, MD was a good reader with a spelling age of over two years below his chronological age. His low score on the Windows test suggested that poor visual recognition skill may have been a factor in his spelling retardation (he spelled using a predominantly phonological strategy). Other individuals were found who corresponded to the "Chinese" subtype of poor speller; CC (Study 3) and GY (Study 7) were both found to have low scores on tests involving verbal information. HW (Study 3) fulfilled the criteria for Boder's (1973) "mixed" subgroup, with low scores on both Animals and Triangles tests.

However, allocating poor spellers to "Chinese" and "Phoenician" groups on the basis of spelling error alone did not reveal any significant overall differences in cognitive ability. The allocation of subjects into these subgroups was not a contentious exercise; independent observers were found to agree with the author in terms of which subgroups subjects should be allocated to. Therefore, identifying subgroups on the basis of error analysis appears to have some credibility. However, those subgroups were not found to differ on the tests employed in the three studies.

It may be difficult to perform this type of analysis if, as Treiman (1984) suggests, poor spellers lie on a continuum with "Chinese" spellers at one end and "Phoenician" spellers at the other. However, it may be that the distinction between the two types is quantitative rather than qualitative; Chinese spellers may actually be *worse* spellers than Phoenicians. This was certainly the case in Study 3, where the Phoenicians not only spelled more words correctly from the Boder lists than the Chinese, but were also found to have significantly higher IQ scores. In Study 8, however, subjects were matched for spelling ability and so such an analysis was unable to be carried out.

On the basis of the findings in this project, it could be argued that the subtype theory of spelling disability may not be very useful in terms of cognitive theories of dyslexia. Spelling is affected by numerous cognitive variables, and individuals have been found who have deficits at different levels of visual memory, which makes the concept of Boder's (1973) "dyseidetic" subgroup seem over-simplistic.

9.4 Evaluation of the original measures used in the project

In this section, each of the original measures used in the project will be evaluated, with a view to their design and success as psychological instruments, and possibilities for future development.

9.4.1 The Animals test

This test proved to be the only measure which consistently favoured control subjects over poor spellers; it seems therefore to tap a skill which is neither visual memory nor verbal labelling, but perhaps the ability to combine the two. This ability would seem, therefore, to be related in some way to spelling ability. In section 9.5, the implications of these results will be discussed in depth.

The Animals test was originally conceived as a visual sequential memory test in which verbal labelling could be used as a facilitative strategy. Studies 1 and 2 indicated that it was successful in achieving this aim.

9.4.2 The Triangles test

The Triangles test was intended as a measure of visual sequential memory that could not be easily mediated by verbal labelling. In Studies 1 and 2, only a small minority of subjects reported using verbal labels to aid performance; other subjects attempted to use a verbal labelling strategy and abandoned it when it was found to be ineffective. In these studies, this test produced significantly lower scores than the Animals test. Therefore it can be argued that it is primarily a test of visual sequential memory, largely unmediated by verbal factors.

The most significant feature of the Triangles test is that it appears to have removed the need for an experimenter to use an articulatory suppression technique. In Chapter 2 (section 2.2.4) it was argued that articulatory suppression - a technique in which an experimental subject indulges in irrelevant articulation while performing a task - was a laboratory-bound means of suppressing verbal labelling. The creation of the Triangles test was an attempt to devise stimuli that would in themselves suppress labelling and thus make the test both mobile and ecologically valid. It was hoped that such a test might have some use as a diagnostic instrument in educational psychology, and that removing the need for articulatory suppression would eliminate a potentially confounding variable. In this respect the Triangles test appears to have been successful.

9.4.3 The Windows test

The Windows test was a test of visual recognition using the yes/no technique. It was hoped that a test could be created which tapped the skill a speller might use in order to identify that the word written on paper was spelled correctly, which would require that an internal representation of the word was retained accurately. Therefore a "forced choice alternative" technique, where the subject has several options to choose from, seemed to be unsuitable.

As with the visual sequential memory tests, the question of verbal labelling was addressed, and stimuli were chosen that subjects would find relatively difficult to label. A number of the initial items had distinguishing features which pilot subjects reported using labels to remember; these items were removed and in Study 4, verbal labelling appeared to have no facilitative effect on recognition accuracy.

9.4.4 The Pictures test

The Pictures test was created in an attempt to isolate the verbal labelling component of the Animals test that had favoured controls over poor spellers in Study 3. It was hypothesised that poor spellers may take longer to retrieve object names than good spellers and that this slow speed of lexical retrieval was hampering their ability to use a verbal labelling strategy. Furthermore, there seemed to be substantial empirical evidence to suggest that naming speed was a good predictor of reading (and spelling) ability (see section 3.4). Many of these studies (e.g. Denckla and Rudel, 1976a, 1976b) used Rapid Automatised Naming tasks in which subjects were required to read the names of a visually-presented matrix of objects as quickly as possible. However it was felt that a "discrete trial" method of object naming would be more relevant since it tested the rate at which subjects were able to retrieve a specific object name from the internal lexicon.

The Pictures test was successful in that it was a discrete trial test of picture naming that could be run outside the laboratory; Wilson and Cline (1995) argued that such an exercise would be difficult to undertake. However it was not successful in isolating the verbal labelling component of the Animals test; in Study 8, poor spellers did not produce significantly slower naming latencies than controls, and yet their Animals mean was significantly lower. Therefore it is argued that slow speed of lexical retrieval is unlikely to be a factor in spelling disability.

A further point regarding the Pictures test is that picture naming may be entirely unrelated to literacy in general. It may act as a measure of vocabulary, although in this instance the stimuli were selected so that they would be in the vocabulary of 9 year olds, and therefore certainly in the vocabulary of 13 year olds. Huttenlocher and Kubicek (1983) suggest that picture naming involves four distinct cognitive processes - visual processing, activation of the concept, lexical retrieval and articulation. Given the complexity of this activity, it seems unlikely that performance on a naming speed test will correlate with spelling ability (which involves somewhat different processes). As with the Triangles and Windows tests, the Pictures test may identify a processing deficit in an individual which seems to account for his or her spelling disability, but it is unlikely to identify a general deficit common to all poor spellers.

9.5 Theoretical implications of the findings of the Animals test

In the preceding sections, it has been argued that the cognitive processes involved in spelling are highly complex. Only a few individuals will be able to account for their spelling difficulty in terms of, say, a deficit in visual recognition. It may be that it is impossible for one test to tap the interconnection of cognitive skills required for spelling.

Nevertheless, the Animals test favoured controls over poor spellers in both controlled studies in this project. A test of visual sequential memory in which verbal labelling does not appear to be a facilitative strategy (the Triangles) failed to differentiate between poor spellers and controls, so it might appear that verbal labelling is the important factor; however, a test measuring speed of lexical retrieval (the Pictures) failed to differentiate between the groups in Study 7.

In this section, a number of alternative suggestions are made to account for the difficulty which poor spellers display on the Animals test, drawing on various cognitive theories of spelling and memory.

9.5.1 Automaticity and spelling

The "dyslexic automatisation deficit" hypothesis of Nicolson and Fawcett (1990, 1995) was described in section 3.4.1. The hypothesis suggests that the core deficit in dyslexia is the failure to achieve automaticity - the ability to combine cognitive and/or motor subskills and, through practice, achieve smooth "automatic" performance (Shiffrin and Schneider, 1977). With specific reference to spelling, Ormrod and Lounge (1990) found a negative correlation between time-on-task and accuracy in a spelling test, and they argued that this provided evidence for an automatisation deficit in poor spellers.

In Chapter 8, it was suggested that this may be one explanation for the poor spellers' performance on the Animals test. With tests measuring either visual memory (using stimuli which are difficult to label) or verbal labelling (speed of lexical retrieval) as separate processes, poor spellers perform at a level similar to controls. The Animals test is different from these other tests because of the potentially facilitative effect of a verbal labelling strategy. Indeed many subjects essentially "recode" the visual information into verbal information for the purpose of rehearsal. This transformation requires simultaneous operation in both visual and verbal modalities. Poor spellers have difficulty integrating visual and verbal information (Lennox and Siegel, 1994; Goswami, 1992) and hence find verbal labelling less effective than controls. Frith (1985) has suggested that, in order to attain the "orthographic stage" of spelling development, children need to be able to automatise the subskills of phonological and visual processing.

9.5.2 Automaticity or better strategy use?

Cheng (1985) has criticised automaticity theory on the grounds that smooth performance in combining processes develops through the application of strategies, or "restructuring". Rather than simply becoming more skilled at combining processes, one devises "short cuts" which may be new processes in themselves.

In Hicks' (1980) study it appeared to be *strategy use* which differentiated dyslexic subjects from normal readers. Dyslexics only profited from using a labelling strategy when instructed to by the experimenter. There are no explicit instructions for subjects to use verbal labelling when performing the Animals test, although the majority of subjects - including poor spellers - appear to use verbal labelling explicitly.

One explanation for the poor spellers' performance on the Animals test is that, although they do use a verbal labelling strategy, they fail to use it as efficiently as controls. Some recent research in the development of children's use of strategies (Miller, 1990; Bjorklund and Coyle, 1994) suggests that strategy use may not become effective until children are as old as 13 years. Miller (1990) refers to "utilisation deficiency" as reflecting the "*developmental lag between spontaneously producing the strategy and receiving any* *benefits from it*" (p. 160). It may be that the poor spellers in Studies 3 and 7 have failed to reach this stage of cognitive development.

The theory that poor spellers fail to make effective use of strategies is consistent with the findings of Gathercole and Hitch (1993), who argue that the development of verbal rehearsal is related to the development of reading and spelling. It may be that the strategy deficiency experienced by poor spellers is the result of an inability to rehearse the items effectively, suggesting a deficit at the level of the phonological loop (Baddeley, 1990). However, the deficit does not appear to be one of lexical retrieval speed, since poor spellers performed as well as controls on the Pictures test in Study 7.

9.5.3 Constraints on effective strategy use

In the previous two sections, two alternative proposals have been made to explain why the Animals test favoured controls over poor spellers in Studies 3 and 7. The first concerned the possible difficulty of automatising and integrating cognitive subskills; the second concerned the effective use of strategies. In this section, models of memory are discussed which might explain how both these explanations are equally plausible.

If an individual has difficulty automatising subskills, and thus fails to make effective use of strategies, it may reflect a limited processing capacity. Guttentag (1995) has argued that processing capacity demands place a constraint on strategy use with younger children. One study (Guttentag, 1984) involved children memorising a word list at the same time as performing a finger-tapping task; the amount of interference from the finger-tapping task was found to decline with age. This was interpreted as suggesting that older children require less processing capacity to perform dual tasks, and this includes the successful use of strategies. In this respect, strategy use and automaticity are governed by the same attentional system.

Brown and Loosemore (1994), whose connectionist model of spelling development was reviewed in section 1.2.1.4, argue that processing capacity is a major restraint on the development of spelling because of the acquisition of rules and strategies, and the integration of phonological and orthographic information. Spelling development, as in the Frith (1985) model, is seen as the "*task of mastering the statistical associations between a set of patterns representing the phonological forms of words and a set of patterns representing the orthographic forms. Under this characterization, the computational difficulty of the mapping will determine how rapidly it is learned.*" (Brown and Loosemore, 1994, p.333). Lack of "computational resources", or processing capacity, they argue, is the core deficit in developmental dyslexia.

It could therefore be argued that poor spellers have a deficit at the level of the central executive in working memory (Baddeley, 1990). This component of the model is the "processing space" in which the slave systems of the phonological loop and visuo-spatial scratch pad are supervised and co-ordinated. The Animals test would seem to be a good example of this operation since it involves the presentation of visual information and the rehearsal capacity of the phonological loop. Indeed, Morris (1987), in an exploration of the visuospatial scratch pad component of working memory, has argued that the central executive is involved in the *encoding*, though not the maintenance, of visually presented material. Therefore, the resources available in the central executive may determine how effectively one uses a strategy, while the separate subsystems are involved in the visible persistence and phonological rehearsal of that material. Co-ordination of these subsystems could be regarded as automaticity.

Some support for this argument comes from the model of the control of action of Norman and Shallice (1986). This model proposes the existence of a "supervisory activating system" which is responsible for achieving skilled performance, or automaticity, by control and planning of actions. Patients with damage to the frontal lobes have been found to display impairment in the use of strategies (Shallice, 1982), attributed to the breakdown of planning behaviour. Baddeley (1990) has suggested that this system performs a similar function to the central executive of working memory.

It appears, therefore, that a number of cognitive theories of skilled performance arrive at similar conclusions regarding the constraints on strategy use. The poor spellers' performance on the Animals test may result from a retardation of the development of processing space, a suggestion that can be linked to both Frith's (1985) and Brown and Loosemore's (1994) theories of spelling development.

9.6 Suggestions for future research

In this section, a number of theories have been put forward to attempt to explain the findings of the project. Visual memory, it has been suggested, is too complex a process to be assessed by one test alone. If further research in this area was to be profitable it may have to consider the process of visual imagery; how words are represented internally, and how these representations might become activated through an imagery process. One way in which this might be investigated is through the measurement of electrical activity in the visual cortex. Several studies have demonstrated a link between the use of visual imagery and a corresponding increase in activity in this area, as measured by both EEG and cerebral blood flow techniques. It might be possible to study EEG activity in the visual cortex during the spelling process, comparing activity during the spelling of regular words with activity during the spelling of irregular words. (In order to prevent the interference of semantic imagery, abstract words would need to be used.) It might also be possible to observe the precise onset of the imagery and the persistence of the trace. Might a visual image be generated towards the end of the spoken stimulus (i.e. the word that is to be spelled), or only afterwards? Might the trace persist while the subject was writing the word down? Such a study may shed considerable light on the act of spelling itself.

However, the major finding from this project concerns the findings of the Animals test. Poor spellers appear to be at a disadvantage when required to operate in more than one modality, and this may reflect difficulty in using the dual route system of spelling described in Chapter 1. It is suggested that poor spellers experience particular difficulty in combining both visual and phonological information, and that this manifests itself in difficulty from reaching the "orthographic" stage of spelling (Frith, 1985).

Further research in this area might proceed along three separate, though closely connected, lines of enquiry. Firstly, studying the link between automaticity and spelling; secondly, studying the link between strategy effectiveness and spelling; and thirdly, studying other functions governed by the central executive of working memory and any possible links with spelling.

Using the Baddeley & Hitch (1974) working memory model as the paradigm, it could be said that, if the phonological loop component is the essential structure for reading skill, then it is the central executive which is the essential component for spelling. This may be related to Norman and Shallice's (1986) model in which a similar component (the supervisory attentional system) is said to co-ordinate functions which are under the control of the frontal lobes. The frontal lobe area is considered to be the site for *planning* behaviour. It could be argued that planning is an essential feature of spelling though not of reading; in reading one has simply to respond to a visual stimulus, while in spelling one has to generate a response, often without any prompt. The formal spelling test requires cross-modal transfer of information from an auditory stimulus, mediated perhaps by visual memory (via visual imagery), culminating in a motor activity. Therefore, planning is highly important, as is the integration of subskills.

It would seem that the processes of automaticity, and of strategy use, may be under the control of the central executive. A future study in this area might involve administering a selection of tasks to poor spellers (and controls) to determine whether these processes are linked. Such a study might not only prove useful in the field of spelling research, but would benefit considerably research into working memory in general, since the central executive is the least clear component of the model and requires further investigation (see Baddeley, 1996).

A possible task for automaticity - which builds on the idea of the Animals test - might involve subjects recalling place names on an imaginary map. Not only would they be expected to recall the precise locations of the places (engaging visuospatial memory) but they would also be expected to recall the precise spelling of the place names (engaging visual/phonological memory). Such a task might be a thorough test of the ability to operate simultaneously in two modalities, and increments in performance across trials would measure the automatisation of this ability.

Strategy effectiveness might be assessed through the type of tests devised by Miller (e.g., Miller, Woody-Ramsey and Aloise, 1991), where subjects are required to perform a memory task which can be considerably enhanced by effective strategy use.

A third task might investigate the use of memory updating, or "running memory" (Morris and Jones, 1990). This skill is related to strategy use and is regarded as a function of the central executive. One way this might be tested is by asking subjects to identify ten familiar personal telephone numbers which have been faithfully committed to long-term memory, and then ask them memorise a set of new numbers for those people. However, as has been suggested (Morris and Jones, 1990), this task might be too dependent on the operation of the phonological loop, so a task tapping this component (such as a digit span measure) would need to be used as a control.

It is argued that combined performance on these three tasks - map learning, strategy use and updating memory - would reflect the operation of the central executive, and if poor spellers performed at a significantly lower level than controls, it would indicate that this component of working memory is closely associated with spelling.

9.7 Concluding remarks

The findings of this project can be summarised as follows:

- Visual memory, as tested by a selection of reliable measures designed to counteract verbal labelling wherever possible, does not appear to discriminate between poor spellers and competent spellers at age 13. At an individual level, poor visual memory may be identified as a primary cause of spelling difficulty, but this does not generalise to the overall poor spelling population.

- A test requiring subjects to integrate both verbal and visual information (the Animals test) was found to discriminate between poor spellers and competent spellers, with the competent spellers attaining significantly higher test scores.

- On a test of picture naming, in which subjects are required to name objects presented in single trials, there was no significant difference in response latency between poor spellers and competent spellers.

- From this set of results it is argued that poor spellers have a specific deficit in their ability to integrate information when required to operate within more than one modality. This finding is consistent with theories of developmental dyslexia, and suggests that the level of working memory which is important in spelling is the central executive, a putative structure which controls, plans and integrates other cognitive processes.

REFERENCES

Alegria, J., & Mousty, P. (1994). On the Development of Lexical and Non-Lexical Spelling Procedures of French-Speaking, Normal and Disabled Children. In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Anastasi, A. (1990). *Psychological Testing: Sixth Edition.* New York: MacMillan.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence (Ed.), *The psychology of learning and motivation: advances in research and theory Vol. 2* (pp 89-195). New York: Academic Press.

Backman, J.E., Mamen, M., & Ferguson, H.B. (1984). Reading level design: Conceptual and methodological issues in reading research. *Psychological Bulletin*, 96, 560-568.

Baddeley, A.D. (1986). *Working Memory.* Oxford University Press.

Baddeley, A.D. (1990). *Human Memory.* Lawrence Erlbaum Associates: Hove.

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5-28.

Baddeley, A.D., Eldridge, M., & Lewis, V. (1981) The role of subvocalisation in reading. Quarterly *Journal of Experimental Psychology*, 33A, 439-454.

Baddeley, A.D., Emslie, H., & Nimmo-Smith, I. (1995). *Doors and People: A test of visual and verbal recall and recognition.* Bury St Edmunds: Thames Valley Test Company.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation, Vol. 8*. New York: Academic Press.

Baddeley, A.D., & Lieberman, K. (1980). Spatial working memory. In R. Nickerson (Ed.) *Attention and Performance*. Hillsdale, N.J: Lawrence Erlbaum Associates.

Baddeley, A.D., Lewis, V.J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36, 233-252.

Baddeley, A.D., & Scott, D. (1971). Short-term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology*, 23, 275-283.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 14, 575-589.

Baddeley, A.D., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, 24, 490-502.

Badian, N.A. (1994). Do dyslexic and other poor readers differ in reading-related cognitive skills? *Reading And Writing*, 6, 45-63.

Badian, N.A. (1993). Phonemic awareness, naming, visual symbol processing and reading. *Reading And Writing*, 5, 87-100.

Bahrick, H.P., & Boucher, B. (1968). Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 78(3), 417-422.

Barry, C. (1994). Spelling Routes (or Roots or Rutes). In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Batchelor, E.S., Kixmiller, J.S., & Dean, R.S. (1990). Neuropsychological aspects of reading and spelling performance in children with learning disabilities. *Developmental Neuropsychology* 6, 183-192.

Bear, D.R., & Barone, D. (1991). The relationship between rapid automatized naming and orthographic knowledge. *National Reading Conference Yearbook*, 40, 179-184.

Bell, T.E. (1990). Rapid sequential processing in dyslexic and ordinary readers. *Perceptual and Motor Skills*, 71, 1155-1159.

Betts, G.H. (1909). *The distribution and functions of mental imagery*. New York: Columbia University Teachers College, Contributions to Education Series 26.

Bishop, D.V.M., & Robson, J. (1989). Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of "Articulatory Coding". *Quarterly Journal of Experimental Psychology*, 41A, 123-140.

Bjorklund, D.F., & Coyle, T.R. (1995). Utilization Deficiencies in the Development of Memory Strategies. In F.E.Weinert & W.Schneider (Eds.), *Memory Performance and Competencies*. New Jersey: Lawrence Erlbaum Associates.

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. *Developmental Medicine and Child Neurology*, 15, 663-687.

Boder, S., & Jarrico, S. (1982).*The Boder test of reading-spelling patterns*. New York: Grune & Stratton Inc.

Bowers, P.G., & Swanson, L.B. (1991). Naming speed deficits in reading disability: multiple measures of a singular process. *Journal of Experimental Child Psychology*, 51, 195-219.

Bradley, L., & Bryant, P.E. (1985). *Rhyme and Reason in Reading and Spelling*. Academy for Research in Learning Disabilities Monograph Series, No. 1. Ann Arbor: University of Michigan.

Bradley, L., & Bryant, P.E. (1981). Visual memory and phonological skills in reading and spelling backwardness. *Psychological Research*, 43, 193-199.

Brannan, J.R., & Williams, M.C. (1988). Developmental versus sensory deficit effects on perceptual processing in the reading disabled. *Perception and Psychophysics* 44, 437-444.

Brown, G.D.A., & Loosemore, R.P.W. (1994). Computational Approaches to Normal and Impaired Spelling. In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Bryant, P.E., & Bradley, L. (1985). *Children's Reading Problems*. Oxford: Blackwell Scientific Publications.

Bryant, P.E., & Bradley, L. (1980). Why children sometimes write words that they do not read. In U. Frith (Ed),*Cognitive Processes in Spelling*. London: Academic Press.

Bryant, P.E., & Goswami, U. (1986). Strengths and weaknesses of the reading level design: A comment on Backman, Mamen and Ferguson. *Psychological Bulletin*, 100, 101-103.

Bub, D.N., & Kertesz, A. (1982). Deep agraphia. *Brain and Language*, 17, 147-166.

Burden, V., & Campbell, R. (1994). The development of word-coding skills in the born deaf: an experimental study of deaf school-leavers. *British Journal of Developmental Psychology*, 12, 331-349.

Campbell, R. (1983). Writing nonwords to dictation. *Brain and Language* 19, 153-178.

Campbell, R. (1985). When children write nonwords to dictation. *Journal of Experimental Child Psychology*, 40, 133-151.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: evidence from a case of acquired dysgraphia. *Cognition*, 26, 59-85.

Carmichael, L., Hogan, H.P., & Walter, A.A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 15, 73-86.

Ceci, S.J., Leichtman, M.D., & Bruck, M. (1995). The suggestibility of children's eyewitness reports: Methodological issues. In F.E.Weinert & W.Schneider (Eds.), *Memory Performance and Competencies*. New Jersey: Lawrence Erlbaum Associates.

Cheng, P.N. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, 92, 414-423.

Clark, H.J. (1965). Recognition memory for random shapes as a function of complexity, association value, and delay. *Journal of Experimental Psychology*, 69, 590-595.

Cleaves, W.T. (1977). Comparisons of reaction time patterns in a sequential visual recognition task with simple geometric forms. *Perception and Psychophysics*, 22, 191-200.

Cohen, N.J., & Squire, L.R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia using perceptual learning. *Cortex*, 17, 273-278.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, 27, 183-228.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55, 75-84.

Conrad, R. (1972). Speech and Reading. In J.F. Kavanagh & J.G. Mattingly (Eds.),*Language by Ear and by Eye*. MIT Press.

Cornelissen, P., Bradley, L., Fowler, S., & Stein, J. (1994). What children see affects how they spell. *Developmental Medicine and Child Neurology* 36, 716-727.

Crispin, L., Hamilton, W., & Trickey, G. (1984). The relevance of visual sequential memory to reading. *British Journal of Educational Psychology*, 54, 24-30.

Dale, P.S. (1972). *Language Development: Structure and Function*. Hinsdale, Ill.: Dryden Press.

Day, J.B., & Weddell, K. (1972). Visual and auditory memory in spelling: an exploratory study. *British Journal of Educational Psychology* 42, 33-39.

Denckla, M.B., & Rudel, R. (1974). Rapid 'automatised' naming of pictured objects, colours, letters and numbers by normal children. *Cortex*, 10, 186-202.

Denckla, M.B., & Rudel, R. (1976a). Rapid "automatized" naming (R.A.N.): dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14, 471-479.

Denckla, M.B., & Rudel, R. (1976b). Naming of object drawings by dyslexic and other learning disabled children. *Brain and Language*, 3, 1-15.

Dodd, B. (1980). The spelling abilities of profoundly prelinguistically deaf children. In U. Frith (Ed.),*Cognitive Processes in Spelling*. London: Academic Press.

Done, D.J., & Miles, T.R. (1978). Age of word acquisition in developmental dyslexics as determined by response latencies in a picture naming task in M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds.), *Practical Aspects of Memory*. Academic Press: London.

Doughtie, E.B., Wakefield, J.A., Sampson, R.N., & Alston, H.L. (1974). A statistical test of the theoretical model for the Illinois Test of Psycholinguistic Ability (*sic*). *Journal of Educational Psychology*, 66, 410-415.

Ehri, L. (1991). The development of reading and spelling in children: an overview. In M. Snowling & M. Thomson (Eds.), *Dyslexia: integrating theory and practice*. London: Whurr.

Ehri, L. (1980). The development of orthographic images. In U. Frith (Ed.),*Cognitive Processes in Spelling*. London: Academic Press.

Elliott, C.D. (1992). *The British Ability Scales: Handbook and Technical Manual*. Windsor: NFER-Nelson.

Ellis, A.W. (1984). *Reading, Writing and Dyslexia*. Hove: Lawrence Erlbaum Associates.

Ellis, A.W., McDougall, S.J.P., & Monk, A.F. (1996). Are dyslexics different? I. A comparison between dyslexics, reading age controls, poor readers and precocious readers. *Dyslexia*, 2, 31-58.

Ellis, A.W., & Young, A.W. (1988). *Human Cognitive Neuropsychology*. Hove: Lawrence Erlbaum Associates.

Ellis, H.C., & Daniel, T.C. (1971). Verbal process in long term stimulus recognition memory. *Journal of Experimental Psychology*, 90, 18-26.

Ellis, N. (1981). Visual and name coding in dyslexic children. *Psychological Research*, 43, 201-218.

Ellis, N.C., & Large, B. (1988). The early stages of reading: a longitudinal study. *Applied Cognitive Psychology*, 2, 47-76.

Eysenck, M.W., & Keane, M.T. (1995). *Cognitive Psychology: A Student's Handbook (3rd Edition)*. Erlbaum (UK) Taylor and Francis.

Farah, M.J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review* 95, 307-317.

Fawcett, A.J., & Nicolson, R.I. (1992). Automatisation deficits in balance for dyslexic children. *Perceptual and Motor Skills*, 75, 507-529.

Felton, R.H., Wood, F.B., Brown, I.S., Campbell, S.K., & Harter, M.R. (1987). Separate verbal memory and naming deficits in attention deficit disorder and reading disability. *Brain and Language*, 31, 171-184.

Flavell, J.H., Beach, D.R., & Chinsky, J.M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, 37, 283-299.

Ford, S., & Silber, K.P. (1994). Working memory in children: a developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology*, 12, 165-175.

Frick, R.W. (1985). Testing visual short-term memory: Simultaneous versus sequential presentations. *Memory and Cognition*, 13, 346-356.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart & J. Marshall (Eds.), *Surface Dyslexia*. Cambridge: Academic Press, Pp. 301-330.

Frith, U. (1980). Unexpected spelling problems. In U. Frith (Ed.),*Cognitive Processes in Spelling*. London: Academic Press.

Gathercole, S.E., & Hitch, G.J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A.F. Collins, S.E. Gathercole, M.A. Conway, & P.E. Morris (Eds.), *Theories of Memory*. Hove: Lawrence Erlbaum Associates. Pp. 189-209.

Gaylord, R.H. (1969). Estimating test reliability from the item-test correlations. *Educational and Psychological Measurement*, 29, 303-304.

Gerber, M.M., & Hall, R.J. (1987). Information processing approaches to studying spelling deficiencies. *Journal of Learning Disabilities*, 20, 34-42.

Glanzer, M., & Cunitz, A.R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behaviour*, 5, 351-360.

Glosser, G., Goodglass, H., & Biber, C. (1989). Assessing visual memory disorders. *Psychological Assessment*, 1(2), 82-91.

Golden, N.E., & Steiner, S.R. (1969). Auditory and visual functions in good and poor readers. *Journal of Learning Disabilities*, 9, 476-481.

Goldstein, A.G., & Chance, J.E. (1971). Visual recognition memory for complex configurations. *Perception and Psychophysics*, 9, 237-241.

Goodman, R.A., & Caramazza, A. (1986). Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes* 1, 263-296.

Goswami, U. (1992). Annotation: Phonological factors in spelling development. *Journal of Child Psychology and Psychiatry,* 33, 967-975.

Goulandris, N.K., & Snowling, M. (1991). Visual memory deficits: a plausible case of developmental dyslexia? Evidence from a single case study. *Cognitive Neuropsychology*, 8, 127-154.

Gould, S.J. (1981). *The Mismeasure of Man.* Penguin.

Goyen, J.D., & Martin, M. (1977). The relation of spelling errors to cognitive variables and word type. *British Journal of Educational Psychology* 47, 268-273.

Guthrie, J.T., & Goldberg, H.K. (1972). Visual sequential memory in reading disability. *Journal of Learning Disabilities*, 5, 41-46.

Guttentag, R.E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology*, 27, 92-106.

Guttentag, R.E. (1995). Mental Effort and Motivation: Influences on Children's Memory Strategy Use. In F.E.Weinert & W.Schneider (Eds.), *Memory Performance and Competencies.* New Jersey: Lawrence Erlbaum Associates.

Haber, R.N., & Standing, L. (1969). Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology*, 21, 43-54.

Hagen, J.W., & Kingsley, P.R. (1968). Labelling effects in short-term memory. *Child Development*, 39, 113-121.

Hagen, J.W., Meacham, J.A., & Mesibov, G. (1970). Verbal labelling, rehearsal and short-term memory. *Cognitive Psychology* 1, 47-58.

Hanson, V.L., Shankweiler, D., & Fischer, F.W. (1983). Determinants of spelling ability of deaf and hearing adults: Access to linguistic structures. *Cognition* 14, 323-344.

Hatfield, F.M., & Patterson, K.E. (1983). Phonological spelling. *Quarterly Journal of Experimental Psychology*, 35A, 451-468.

Hicks, C. (1980). The ITPA Visual Sequential Memory task: an alternative interpretation and the implications for good and poor readers. *British Journal of Educational Psychology*, 50, 16-25.

Hirshoren, A.A. (1969). A comparison of the predictive validity of the revised Stanford-Binet Intelligence Scale and the Illinois Test of Psycholinguistic Abilities. *Exceptional Children*, 35, 517-521.

Hitch, G.J., Halliday, M.S., Dodd, A., & Littler, J.E. (1989). Development of rehearsal in short-term memory: differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7, 347-362.

Holmes, D.R., & McKeever, W.F. (1979). Material specific serial memory deficit in adolescent dyslexia.*Cortex*, 15, 51-62.

Houghton, G., Glasspool, D.W., & Shallice, T. (1994). Spelling and Serial Recall: Insights from a Competitive Queuing Model. In G.D.A. Brown and N. C. Ellis (Eds.) *Handbook of Spelling*. Chichester: Wiley.

Hulme, C., Silvester, J., Smith, S., & Muir, C. (1986). The effect of word length on memory for pictures: evidence for speech coding in young children. *Journal of Experimental Child Psychology*, 41, 61-75.

Humphreys, G.W., & Bruce, V. (1989). *Visual Cognition*. Hove: Lawrence Erlbaum Associates.

Huston, A.M. (1991). *Understanding Dyslexia*. New York: Madison Books.

Huttenlocher, J., & Kubicek, L.F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 486-496.

Jensen, A.R. (1962). Spelling errors and the serial-position effect. *Journal of Educational Psychology*, 53, 105-109.

Johnson, C.J., & Clark, J.M. (1988). Children's picture naming difficulty and errors: Effects of age of acquisition, uncertainty, and name generality. *Applied Psycholinguistics*, 9, 351-365.

Johnson, D.J., & Myklebust, H.R. (1967).*Learning Disabilities: Education Principles and Practices*. New York: Grune & Statton.

Kaplan, E., Goodglass, H., & Weintraub, S. (1982). *Boston Naming Test*. Philadelphia: Lea & Ferbiger.

Kass, C.E. (1966). Psycholinguistic disabilities of children with reading problems. *Exceptional Children*, 35, 533-539.

Kelly, R.T., & Martin, D.W. (1974). Memory for random shapes: a dual-task analysis. *Journal of Experimental Psychology*, 103, 224-229.

Kirk, S.A., McCarthy, J.J., & Kirk, W.D. (1968).*The Illinois Test of Psycholinguistic Abilities*. Urbana: University of Illinois Press.

Kirk, S.A., & Kirk, W.D. (1971). *Psycholinguistic Learning Disabilities: Diagnosis and Remediation*. Urbana: University of Illinois Press.

Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory and Cognition*, 1, 449-453.

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 47-60.

Lennox, C., & Siegel, L.S. (1994). The Role of Phonological and Orthographic Processes in Learning to Spell. In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Levine, D.N., Calvanio, R., & Popovics, A. (1982). Language in the absence of inner speech. *Neuropsychologia*, 20, 391-409.

Levy, B.A. (1977). Reading: Speech and meaning processes. *Journal of Verbal Learning and Verbal Behaviour*, 16, 623-638.

Leybaert, J. (1992). Determinants of reading ability in deaf individuals: The role of phonological codes. In M. Marschark & D. Clark (Eds.), *Psychological Perspectives on Deafness*. Lawrence Erlbaum Associates.

Link, K., & Caramazza, A. (1994). Orthographic Structure and the Spelling Process: A Comparison of Different Codes. In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31.

Loftus, E.F., & Palmer, J.C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour*, 13, 585-589.

Logie, R.H. (1986). Visuo-spatial processes in working memory. *Quarterly Journal of Experimental Psychology*, 38A, 229-247.

Macione, J.R. (1969). Psychological correlates of reading disability as defined by the Illinois Test of Psycholinguistic Abilities. Doctoral dissertation, University of South Dakota.

Margolin, C.M., Griebel, B., & Wolford, G. (1982). Effect of distraction on reading versus listening. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8, 613-618.

Meyers, C.E. (1969). What the ITPA measures: A synthesis of factor studies of the 1961 edition. *Educational And Psychological Measurement*, 29, 867-876.

Miles, T. (1993). New Developments. In M. Snowling & M. Thomson (Eds.),*Dyslexia: the Pattern of Difficulties*. Whurr.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Miller, P.H. (1990). The development of strategies of selective attention. In D.F. Bjorklund (Ed.), *Children's strategies: Contemporary Views of Cognitive Development.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, P.H., Woody-Ramsey, J., & Aloise, P.A. (1991). The role of strategy effortfulness in strategy effectiveness. *Developmental Psychology*, 27, 738-745.

Milner, B. (1971). Interhemispheric differences and psychological processes. *British Medical Bulletin*, 3, 272-277.

Morris, N. (1987). Exploring the visuo-spatial scratch pad. *Quarterly Journal of Experimental Psychology*, 39A, 409-430.

Morris, N. & Jones, D.M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111-121.

Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.)*Cognitive Processes in Spelling*. London: Academic Press.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78(4), 679-684.

Murray, D.J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 21, 263-276.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Newman, S., Fields, H., & Wright, S. (1993). A developmental study of specific spelling disability. *British Journal of Educational Psychology*, 63, 287-296.

Newton, M.J., & Thomson, M.E. (1982). *The Aston Index (Revised)*. Wisbech, Cambs, UK: Learning Development Aids.

Nickerson, R.S. (1965). Short term memory for complex meaningful configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 19, 155-160.

Nicolson, R.I., & Fawcett, A.J. (1990). Automaticity: A new framework for dyslexia research? *Cognition*, 30, 159-182.

Nicolson, R.I., & Fawcett, A.J. (1995). Dyslexia is more than a phonological disability. *Dyslexia*, 1, 19-36.

Norman, D.A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R.J. Davidson, G.E. Schwarts, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research and Theory*, Vol. 4. New York: Plenum Press.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9, 97-113.

Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273-281.

Olson, A., & Caramazza, A. (1994). Representation and Connectionist Models: The NETSpell Experience. In G.D.A. Brown & N. C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Ormrod, J.E. (1985). Visual memory in a spelling matching task: comparison of good and poor spellers. *Perceptual and Motor Skills*, 61, 183-188.

Ormrod, J.E. & Lounge, J.P. (1990). Automaticity in spelling competence. *Perceptual and Motor Skills*, 71, 384-386.

Osgood, C.E. (1957). A behavioristic analysis of perception and language as cognitive phenomena. In the University of Colorado, *Contemporary Approaches to Cognition*. Cambridge, M.A: Harvard University Press.

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston.

Paraskevopoulos, J.N., & Kirk, S.A. (1969). *The Developmental and Psychometric Characteristics of the Revised Illinois Test of Psycholinguistic Abilities*. Urbana: University of Illinois Press.

Parkin, A.J. (1988). Review of "Working Memory". *Quarterly Journal of Experimental Psychology*, 40A, 187-189.

Perfetti, C.A., Finger, E., & Hogaboam, T. (1978). Sources of vocalization latency differences between skilled and less skilled young readers. *Journal of Educational Psychology,* 70, 730-739.

Phillips, W.A. (1974). On the distinction between sensory storage and short term visual memory. *Perception and Psychophysics*, 16, 283-290.

Phillips, W.A., & Baddeley, A.D. (1971). Reaction time and short-term memory. *Psychonomic Science*, 22, 73-74.

Phillips, W.A., & Christie, D.F.M. (1977). Components of visual memory. Quarterly *Journal of Experimental Psychology*, 29, 117-133.

Postman, L., & Phillips, L.W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 17, 132-138.

Powell, F.C. (1976). *Cambridge Mathematical and Statistical Tables*. Cambridge University Press.

Pumphrey, P.D., & Reason, R. (1992). *Specific Learning Difficulties (Dyslexia)*. Routledge.

Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain. *Psychological Bulletin*, 80, 1-24.

Raven, J.C., Court, J.H., & Raven, J. (1983). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H.K. Lewis.

Rizzo, N.D. (1939). Studies in visual and auditory memory span with special reference to reading disability. *Journal of Experimental Education*, 8, 208-244.

Rudorf, E.H. (1965). The development of an algorithm for American-English spelling. Doctoral dissertation, Stanford University.

Rumelhart, D.E., & Norman, D. (1982). Simulating a skilled typist: a study of skilled motor performance. *Cognitive Science*, 6, 1-36.

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1. Cambridge, MA: MIT Press.

Sakitt, B. (1976). Iconic memory. *Psychological Review*, 83, 257-276.

Salamé, P., & Baddeley, A.D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behaviour*, 21, 150-164.

Shallice, T. (1981). Phonological agraphia and the lexical route in writing. *Brain*, 104, 413-429.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society London B*, 298, 199-209.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge University Press.

Shallice, T., & Warrington, E.K. (1970). Independent functioning of the verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 22, 261-273.

Seidenberg, M.S., & Tannenhaus, M.K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 546-554.

Shepard, R.N. (1978). The mental image. *The American Psychologist*, 33, 125-137.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and general theory. *Psychological Review*, 84, 127-190.

Skurnik, L.S., & Nuttall, D.I. (1969). *Examination and Item Analysis Manual.* Slough: National Foundation for Educational Research.

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, 27, 285-292.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74 (11, Whole No. 498).

Stanovich, K. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.

Stanovich. K.E., Freeman, D.J., & Cunningham, A.E. (1983). The development of the relation between letter-naming speed and reading ability. *Journal of the Psychonomic Society*, 21, 199-202.

Sterling, C.M. (1983). Spelling errors in context. *British Journal Psychology,* 74, 353-364.

Swanson, H.L. (1984). Semantic and visual memory codes in learning disabled readers. *Journal of Experimental Child Psychology*, 37, 124-140.

Swanson, L. (1978). Verbal encoding effects on the visual short-term memory of learning disabled and normal readers. *Journal of Educational Psychology*, 70, 539-544.

Tenney, Y. J. (1980). Visual factors in spelling. In U. Frith (Ed.),*Cognitive Processes in Spelling*. London: Academic Press.

Thomson, M.E. (1984). *Developmental Dyslexia.* Whurr Publishers.

Thomson, M.E., & Newton, M. (1982). *The Aston Index (Revised).* LDA.

Torgesen, J.K. (1985). Memory processes in reading disabled children. *Journal of Learning Disabilities*, 18, 350-357.

Torgesen, J.K., & Goldman, T. (1977). Verbal rehearsal and short-term memory in reading disabled children. *Child Development* 48, 56-60.

Treiman, R. (1984). Individual differences among children in spelling and reading styles. *Journal of Experimental Child Psychology*, 37, 463-477.

Turner, M. (1994). Streamlining Raven. *Dyslexia Review*, 24, 20-22.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science* 185, 1124-1131.

Tzeng, O.J.L. (1971). Positive recency effects in delayed free recall. *Journal of Verbal Learning and Verbal Behaviour*, 12, 436-439.

Vallar, G., & Cappa, S.F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Psychology*, 4, 55-78.

Vanderplas, J.M., & Garvin, E.A. (1959). The association value of random shapes. *Journal of Experimental Psychology*, 57, 147-154.

Vellutino, F.R., Steger, J.A., DeSotto, L., & Phillips, F. (1975). Immediate and delayed recognition of visual stimuli in poor and normal readers. *Journal of Experimental Child Psychology*, 19, 223-232.

Véronis, J. (1988). From sound to spelling in French: simulation on a computer. *Cahiers de Psychologie Cognitive*, 8, 315-334.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *Quarterly Journal of Experimental Psychology*, 48A, 822-848.

Vygotsky, L.S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Walker, B.S. (1974). Vividness of imagery and spelling errors.

Walsh, D.J., Price, G.G., & Gillingham, M.G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly*, 23, 108-122.

Waters, G.S., Bruck, M., & Malus-Abramowitz, M. (1988). The role of linguistic and visual information in spelling: a developmental study. *Journal of Experimental Child Psychology*, 45, 400-421.

Weekes, B.S. (1994). Spelling skills of lexical readers. *British Journal of Psychology*, 8, 245-257.

Wilcox, B.M. (1969). Visual preferences of human infants for representations of the human face. *Journal of Experimental Child Psychology*, 7, 10-20.

Willows, D.M., & Scott, R. (1994). Spelling Processes of the Reading Disabled. In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of Spelling*. Chichester: Wiley.

Wilson, J., & Cline, T. (1995). The naming speed test. *Educational and Child Psychology*, 12, 39-45.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, 14, 1-33.

Wing, A.M., & Baddeley, A.D. (1980). Spelling errors in handwriting: a corpus and a distributional analysis. In U. Frith (Ed.), *Cognitive Processes in Spelling*. London: Academic Press.

Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly*, 26, 123-141.

Wolff, P.H., Michel, G.F., & Ovrut, M. (1990). Rate variables and automatized naming in developmental dyslexia. *Brain and Language*, 39, 556-575.

APPENDIX 1

Program and materials for Animals/Triangles tests

Figure A1. Screen dump of Authorware "flowchart" display for the Animals tests

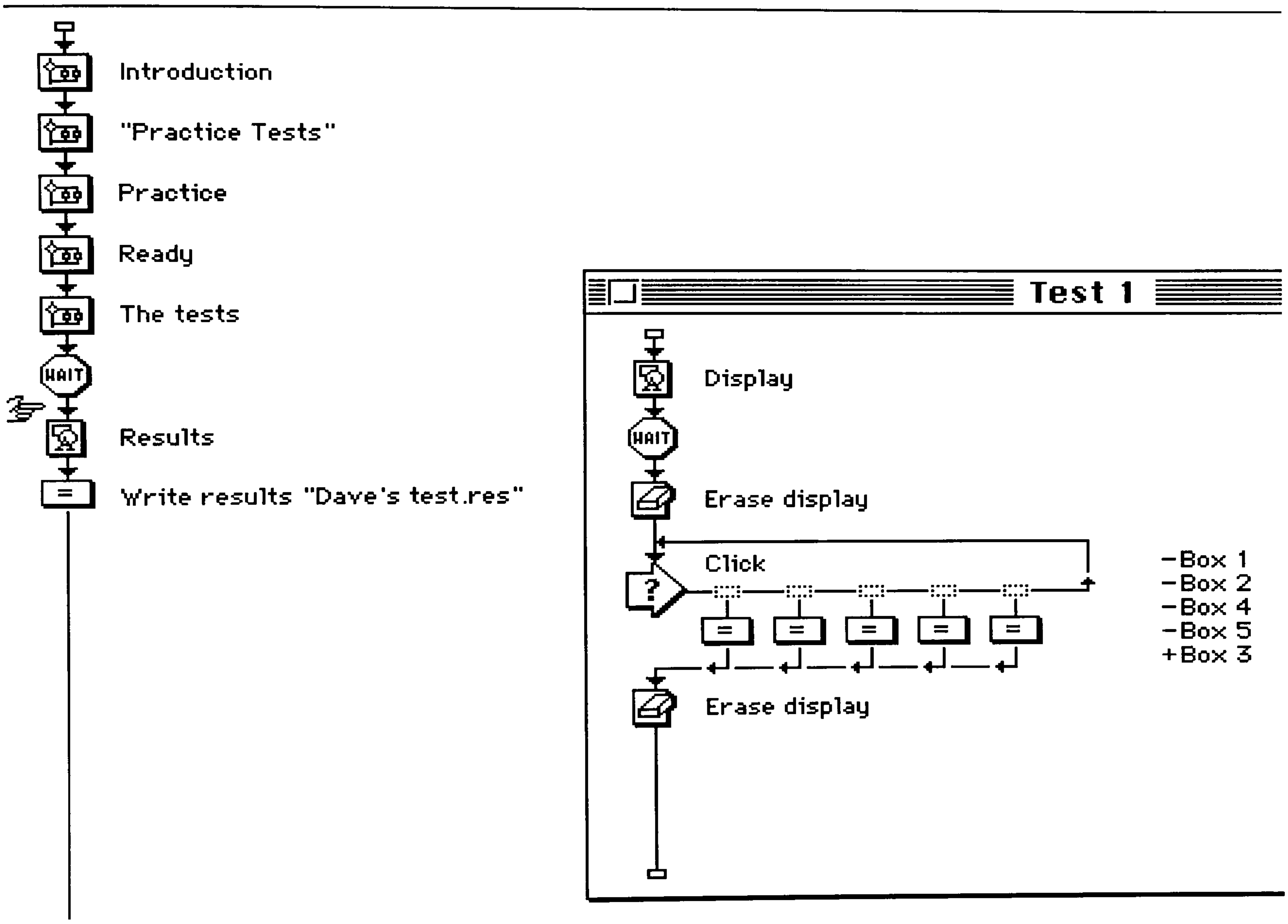


Figure A2. Complete set of stimuli used in Test 1 (the Kirk test)

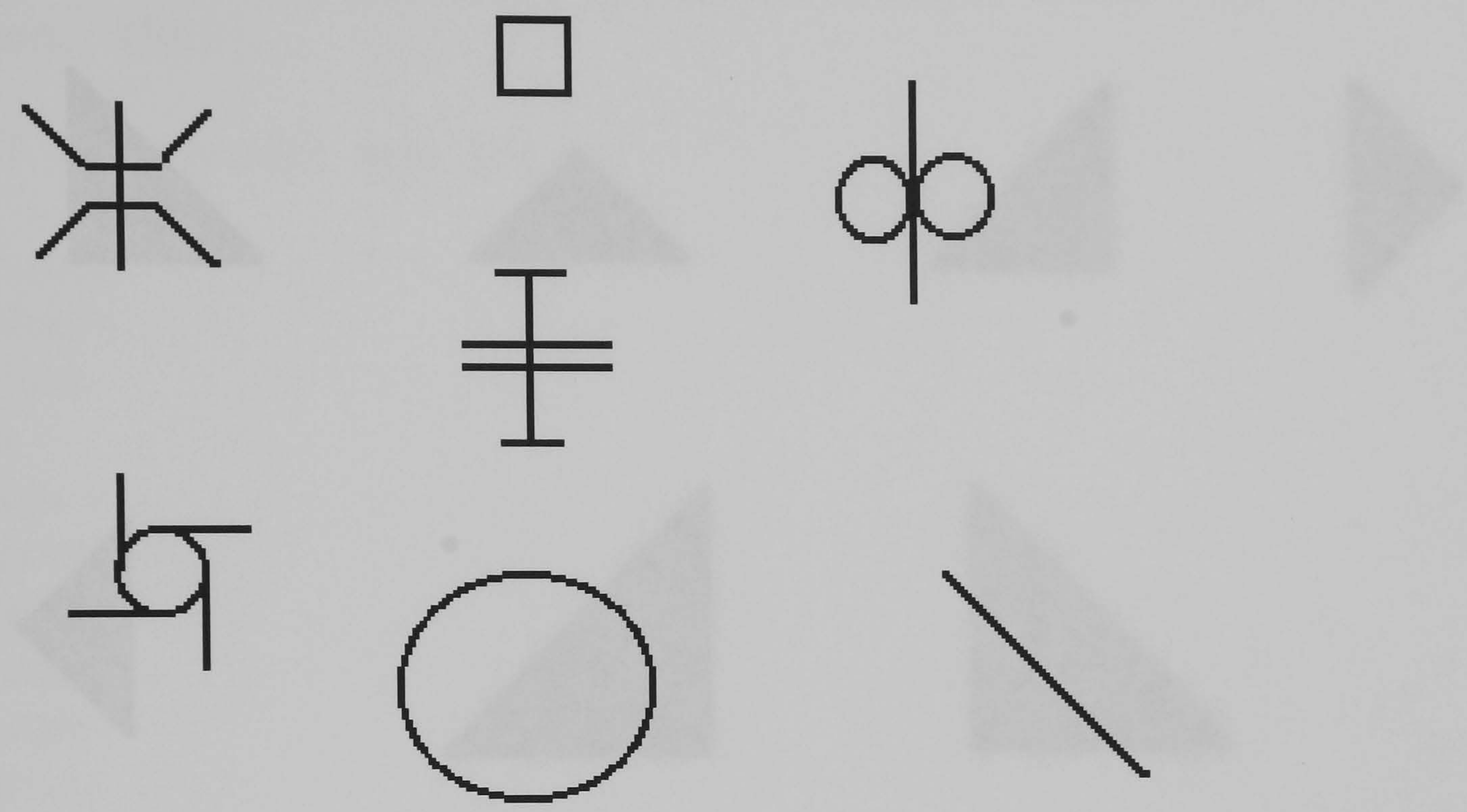


Figure A3. Complete set of stimuli used in Test 2 (Animals test)

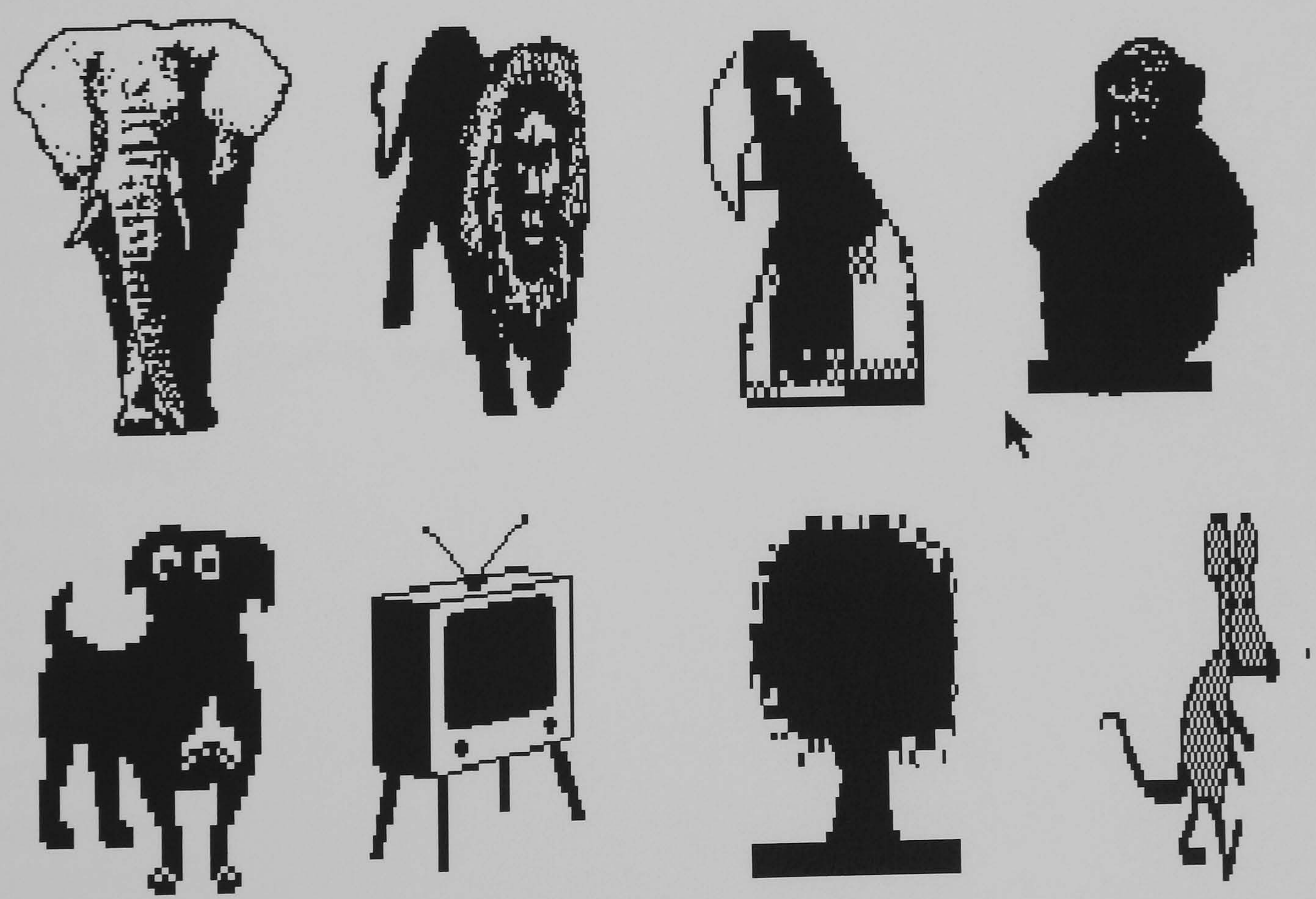
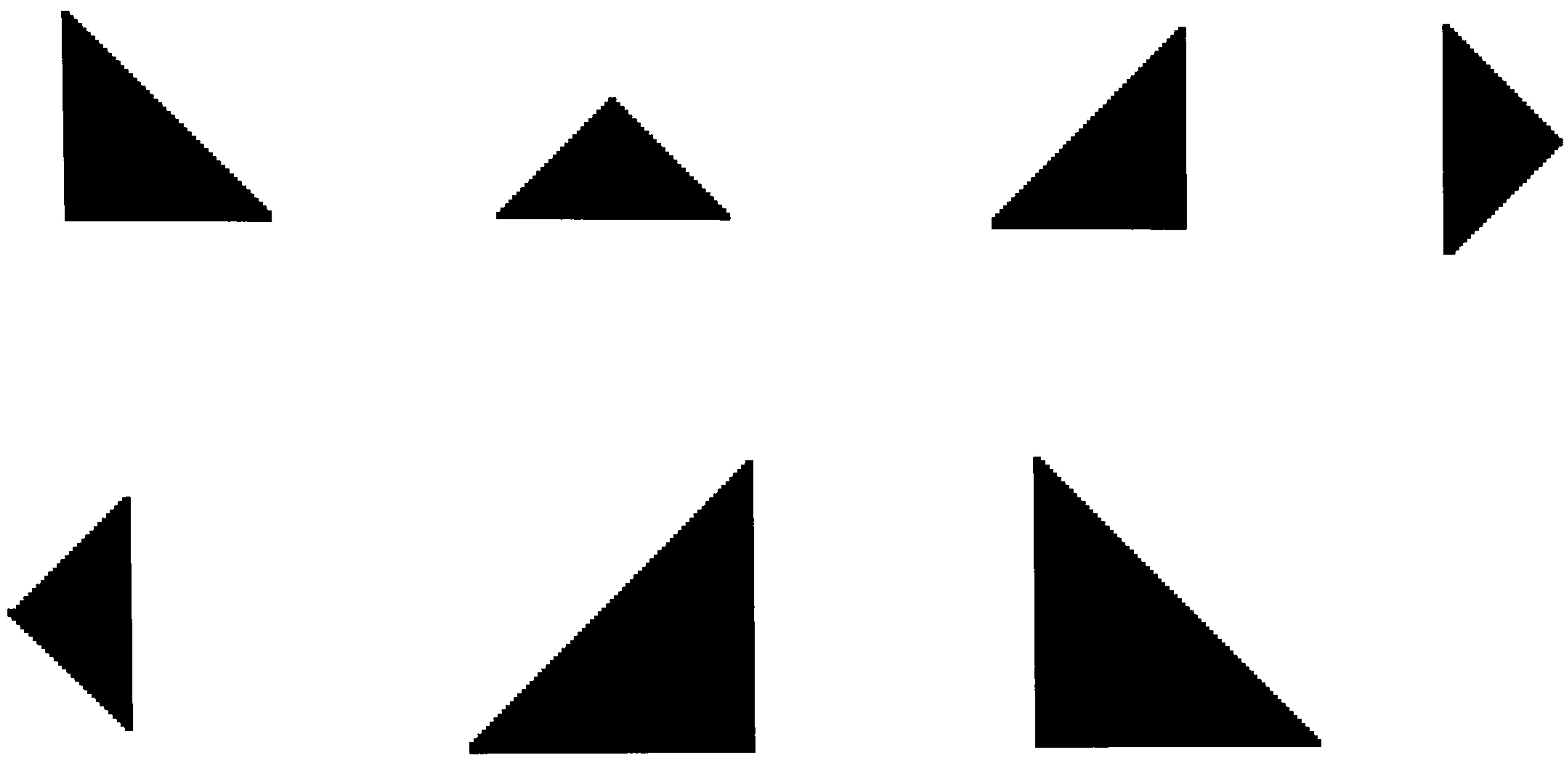


Figure A4. Complete set of stimuli used in Test 3 (Triangles)



APPENDIX 1A

The graded word lists used in Study 3 (from Boder and Jarrico, 1982).

List 7 (7th grade; age 13)

astronomy
doubtful
democrat
hasten
frequent
judgment
quotation
knapsack
publisher
liquor
charity
acknowledge
handicap
cruise
nevertheless
scientific
representative
sergeant
revenge
thorough

List 8 (8th grade; age 14)

discriminate
circuit
fantastic
guarantee
hibernation
lieutenant
perforated
mortgage
unemployed
schedule
armament
acreage
diploma
nourishing
detestable

pursuit
omitted
reigned
testify
temperate

List 9 (9th grade; age 15)

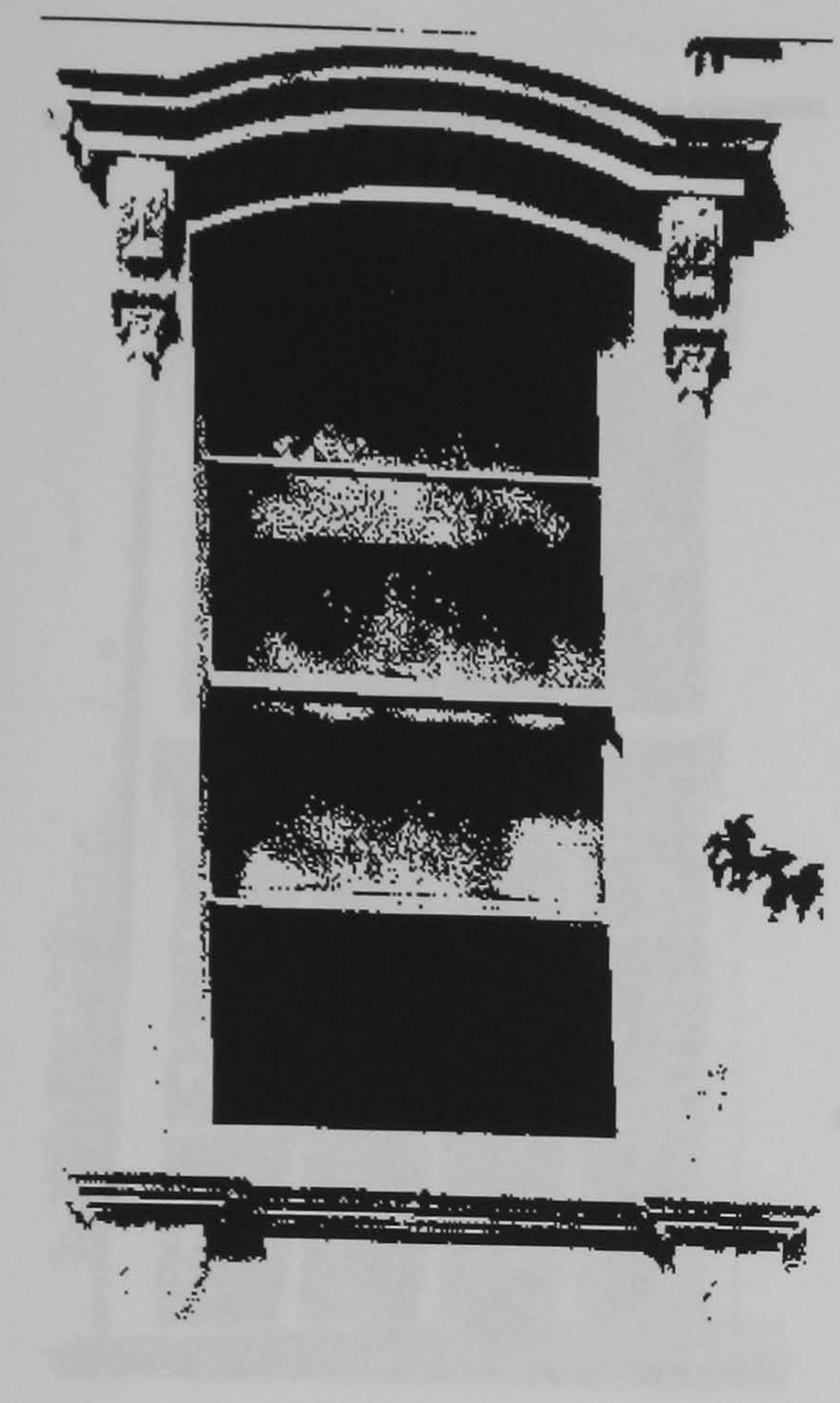
abstract
catastrophe
destitute
chaos
misconduct
encore
optimistic
geyser
remedial
sovereign
insignificant
bosom
prehistoric
champagne
sublime
espionage
telescope
limousine
verify
righteous

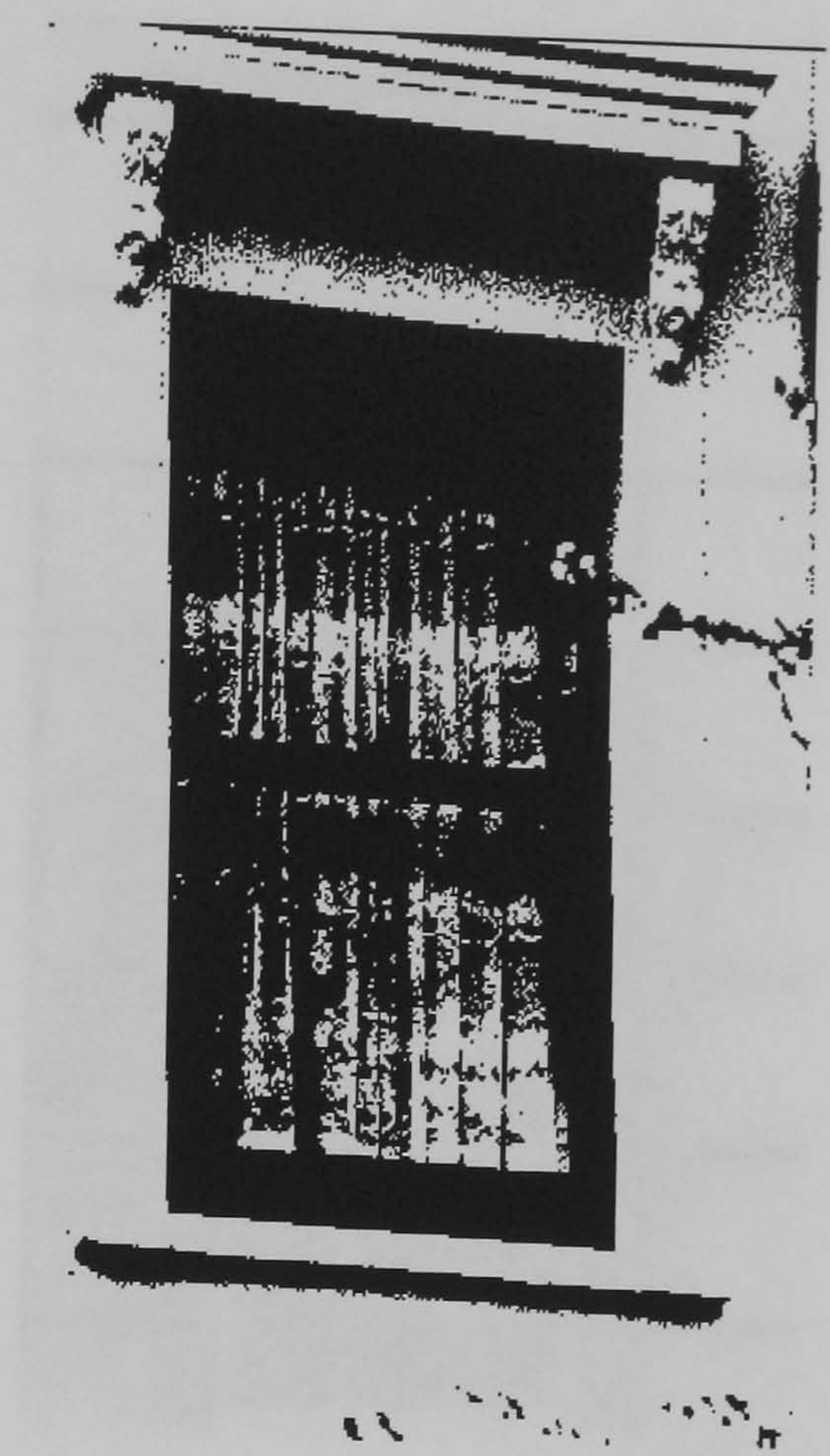
APPENDIX 2

Figure A2. Selected items from the Windows test

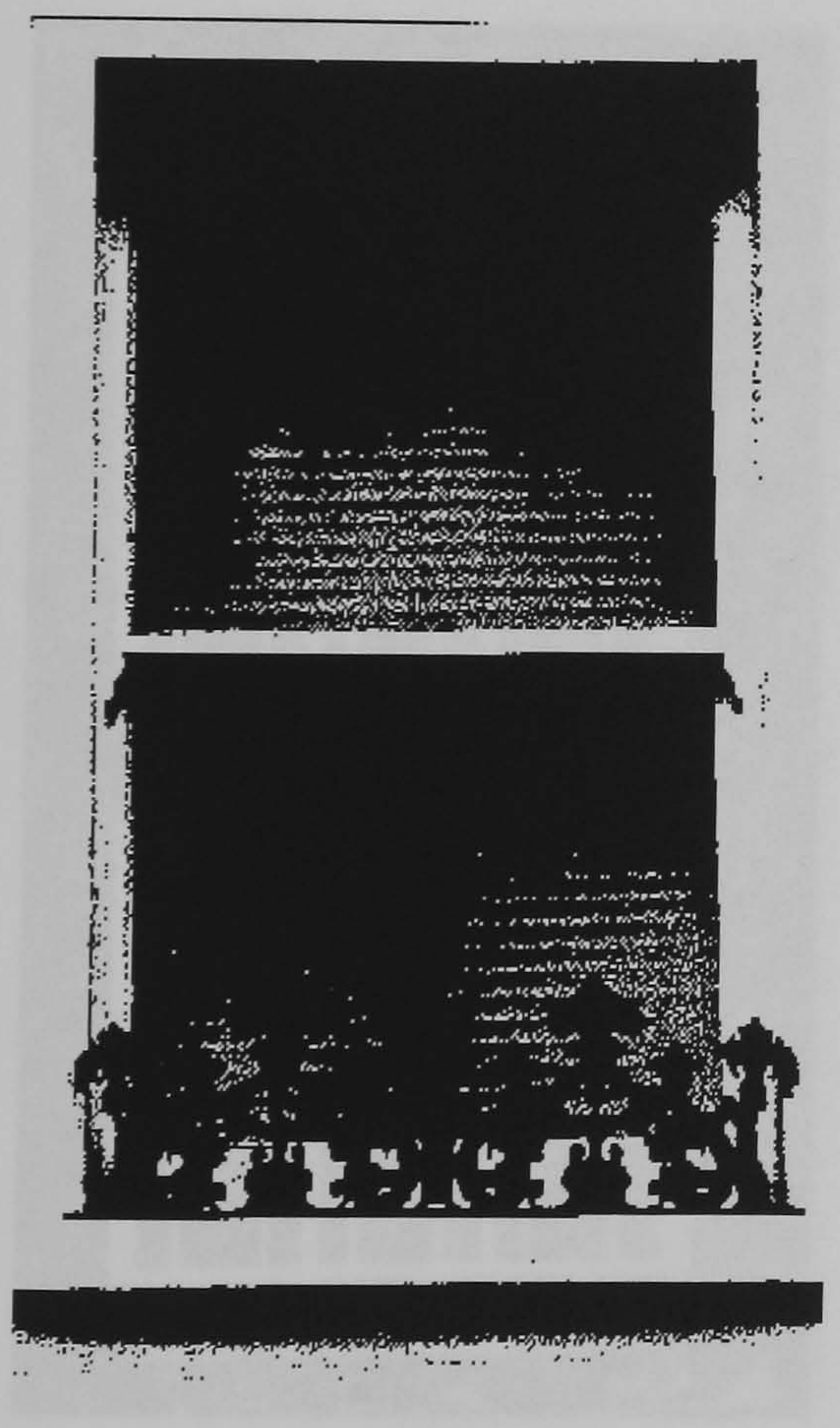
On the next pages, a selection of "shared feature" items and their respective distractors are displayed.

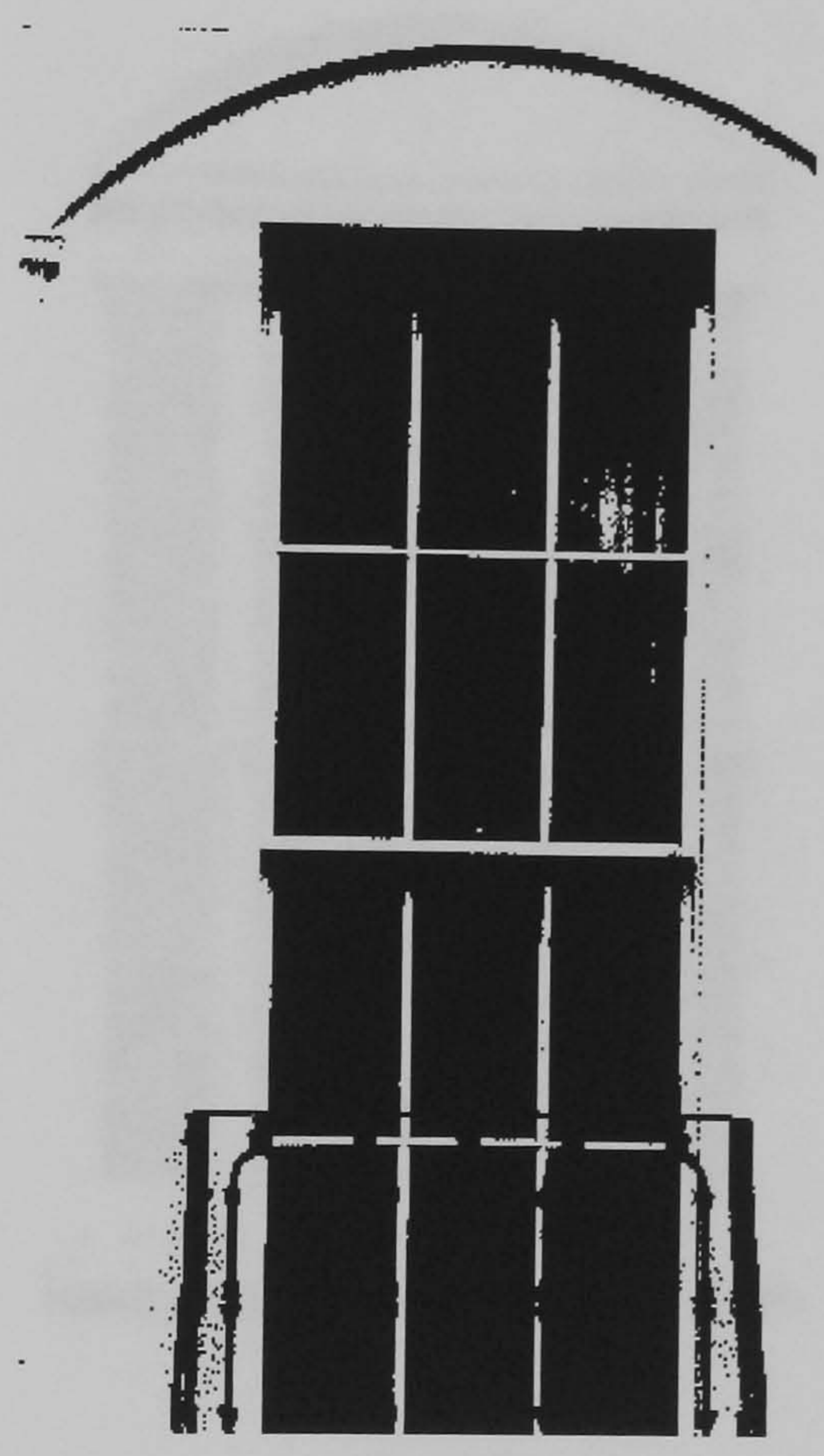
The shared feature for these windows is the arch shape (thus the label "arch" might be used)



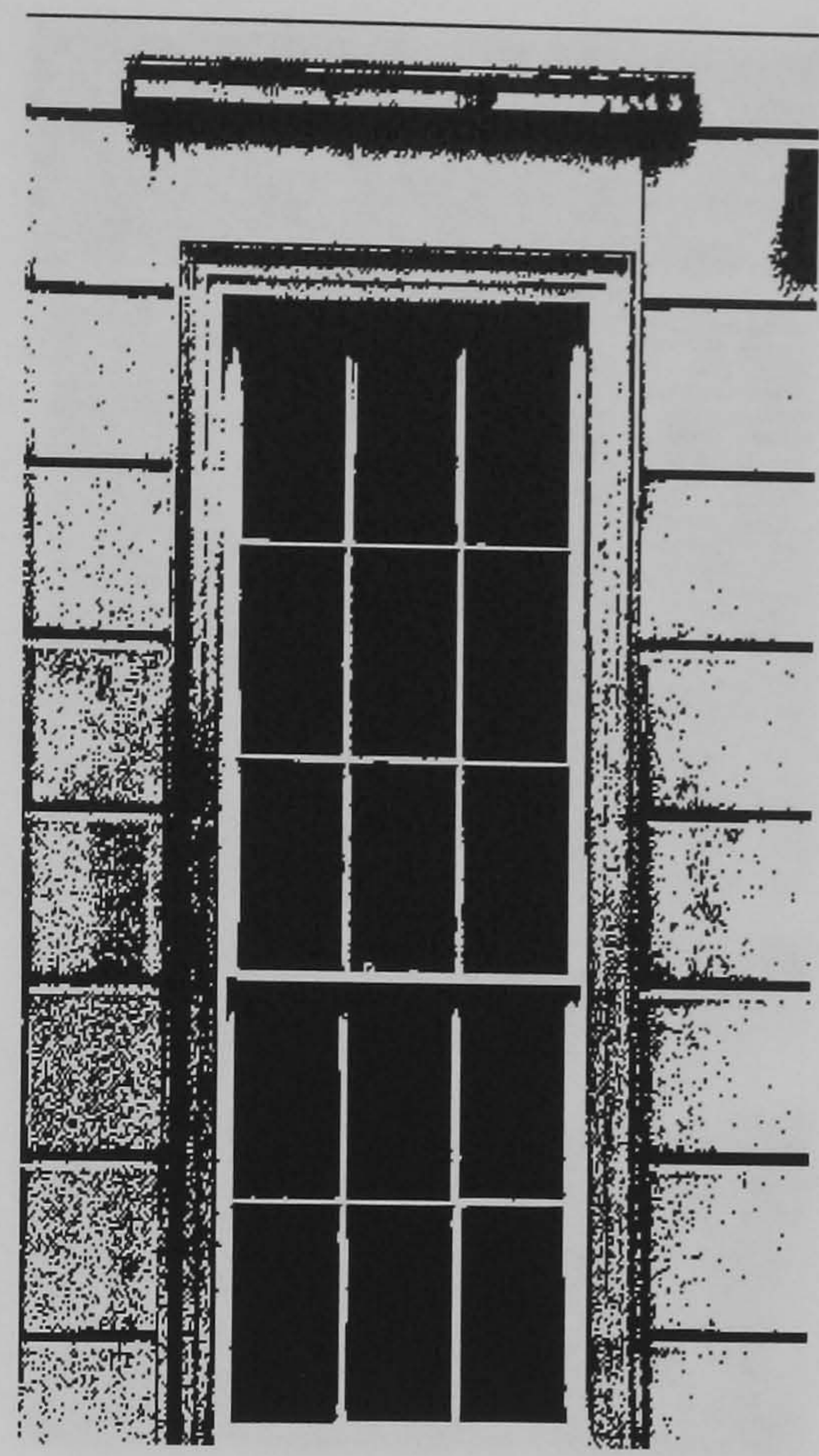


Both these windows have a cornice over the top.

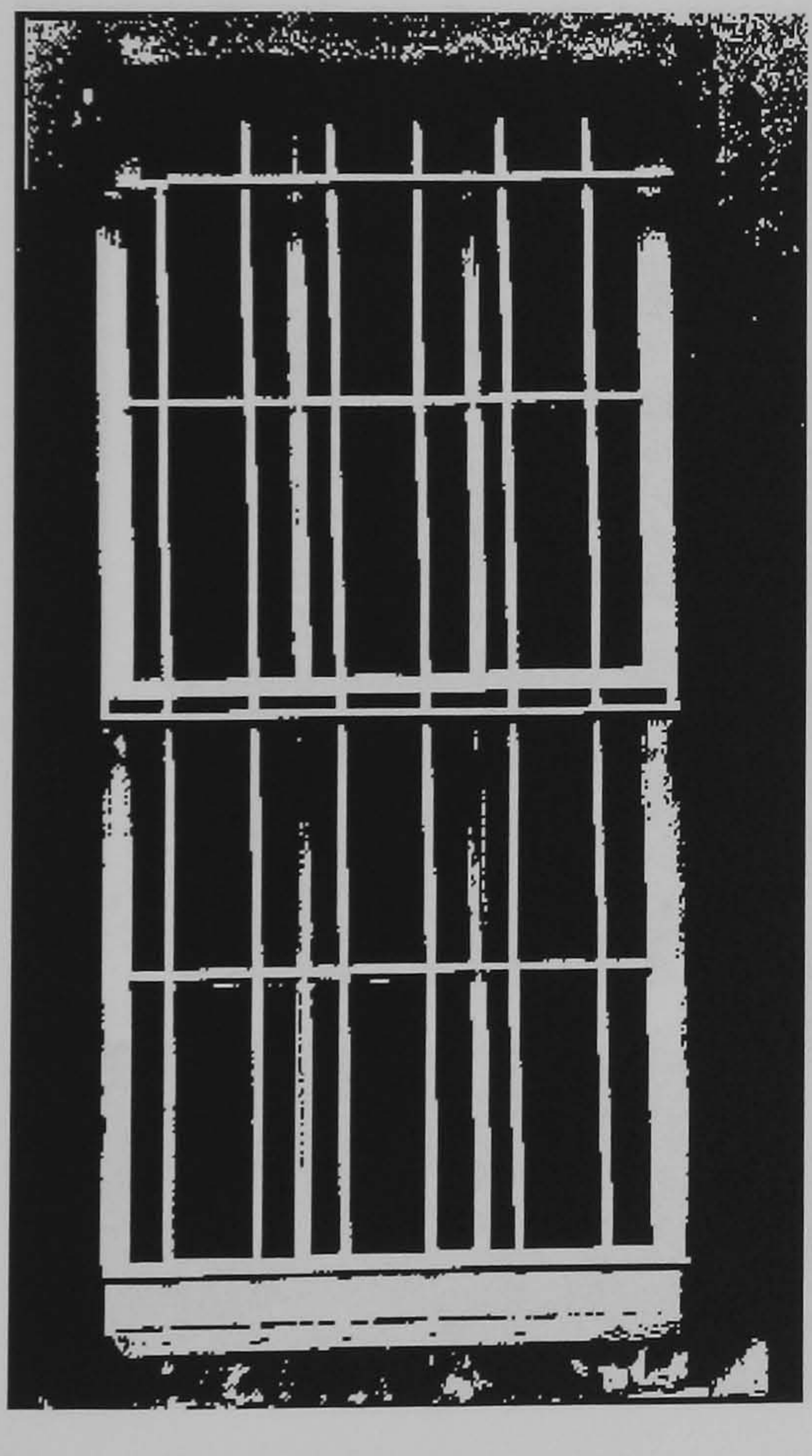


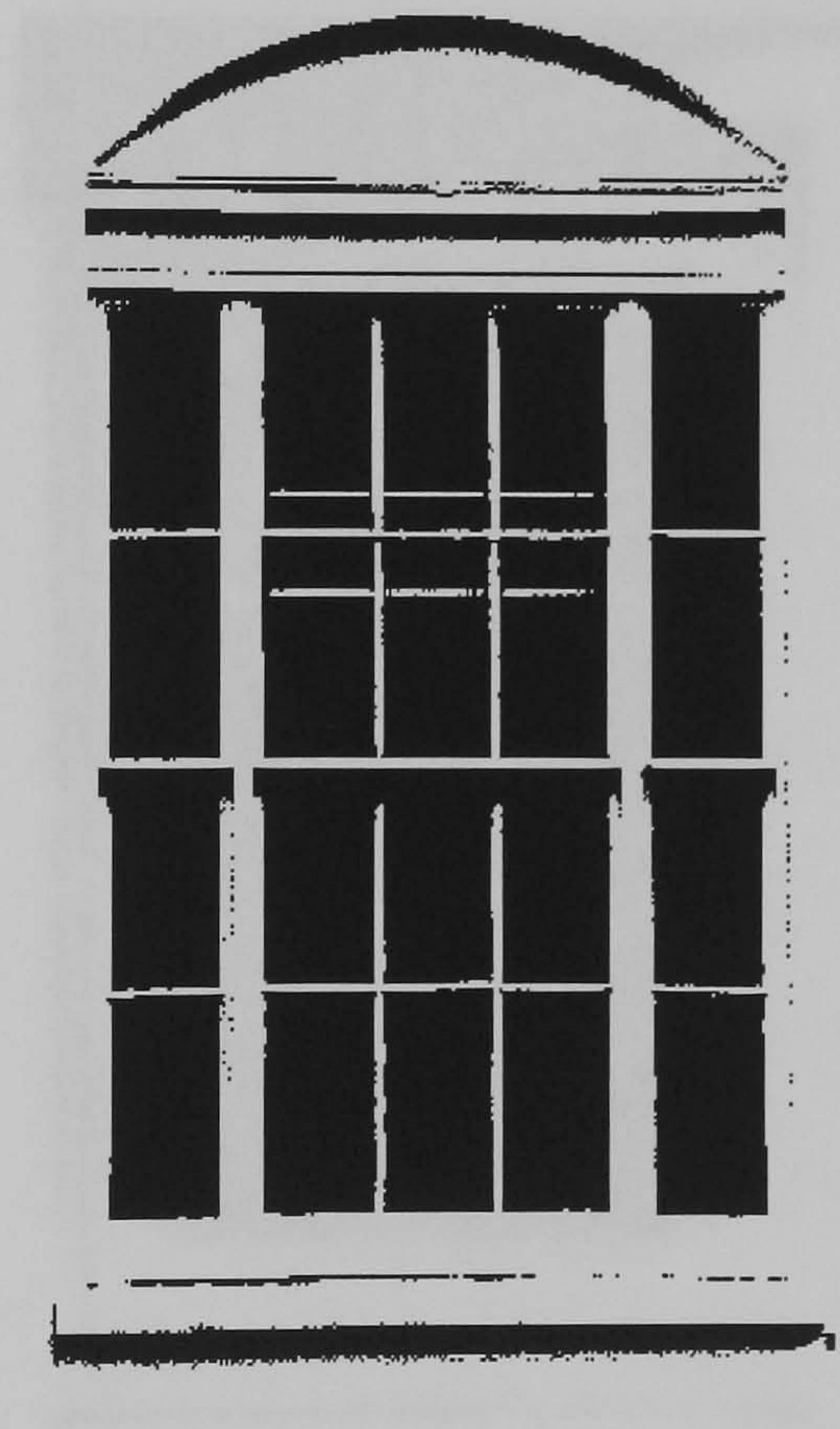


These windows both have black railings at the bottom.

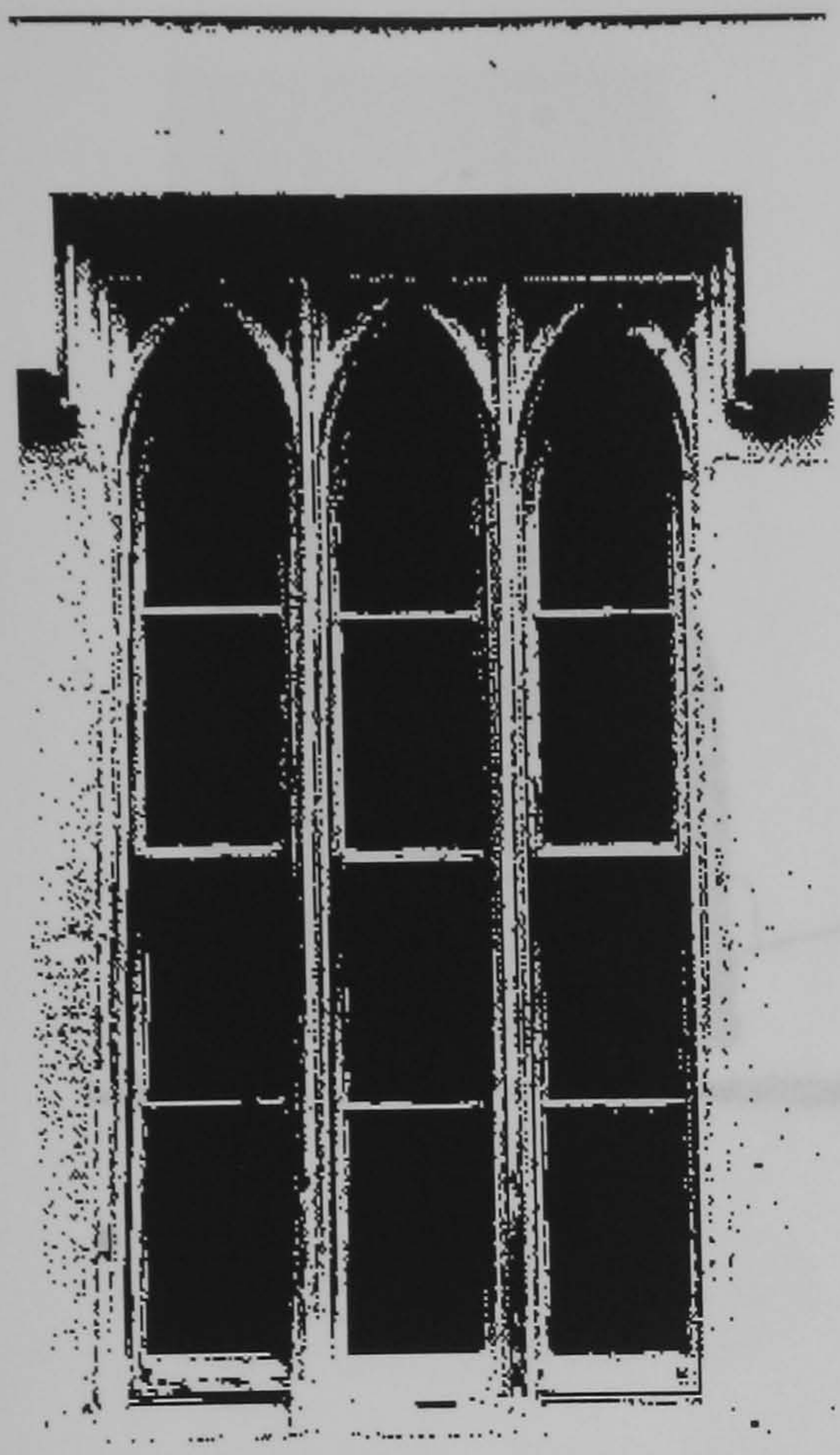


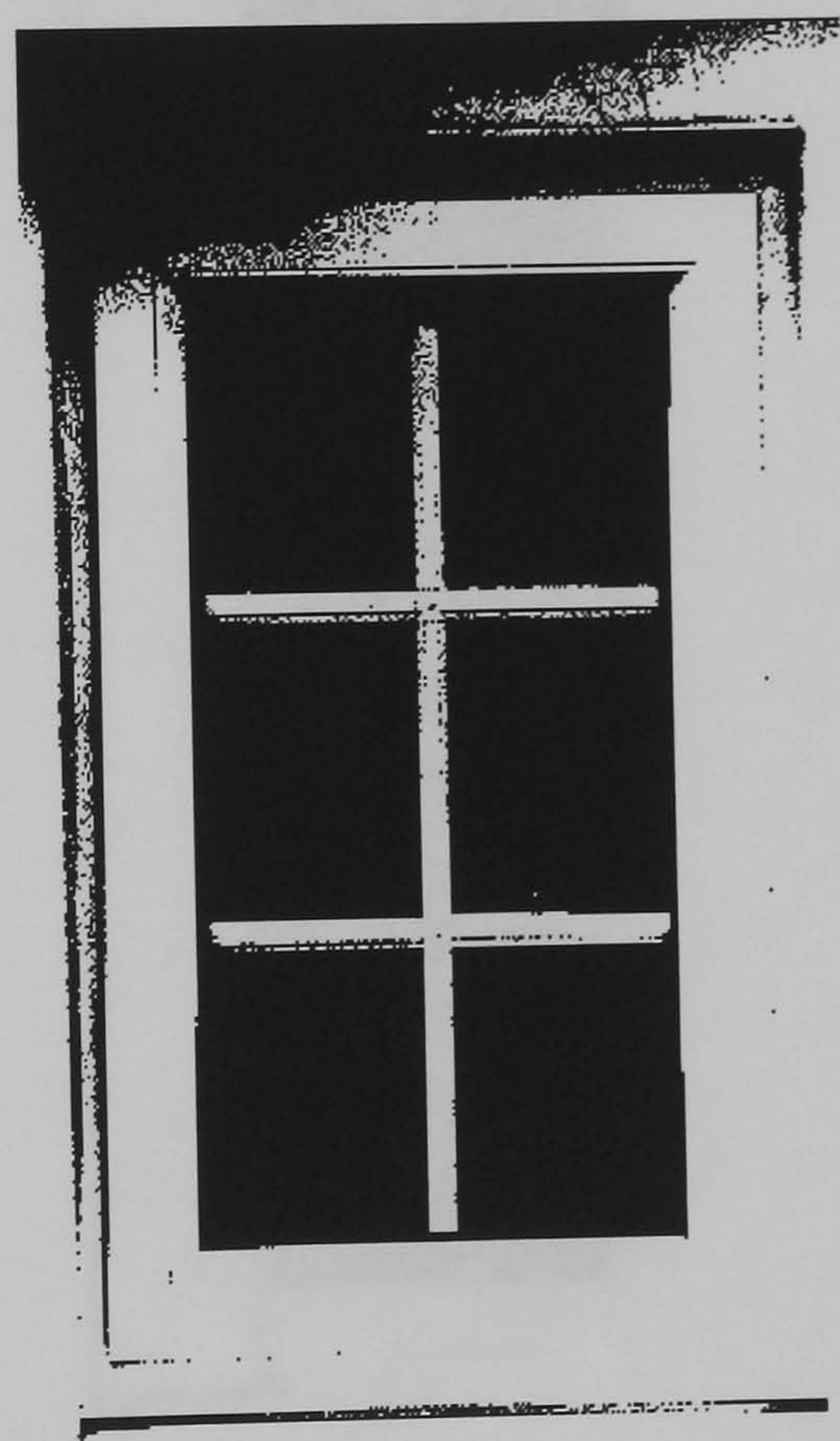
Two multi-pane windows

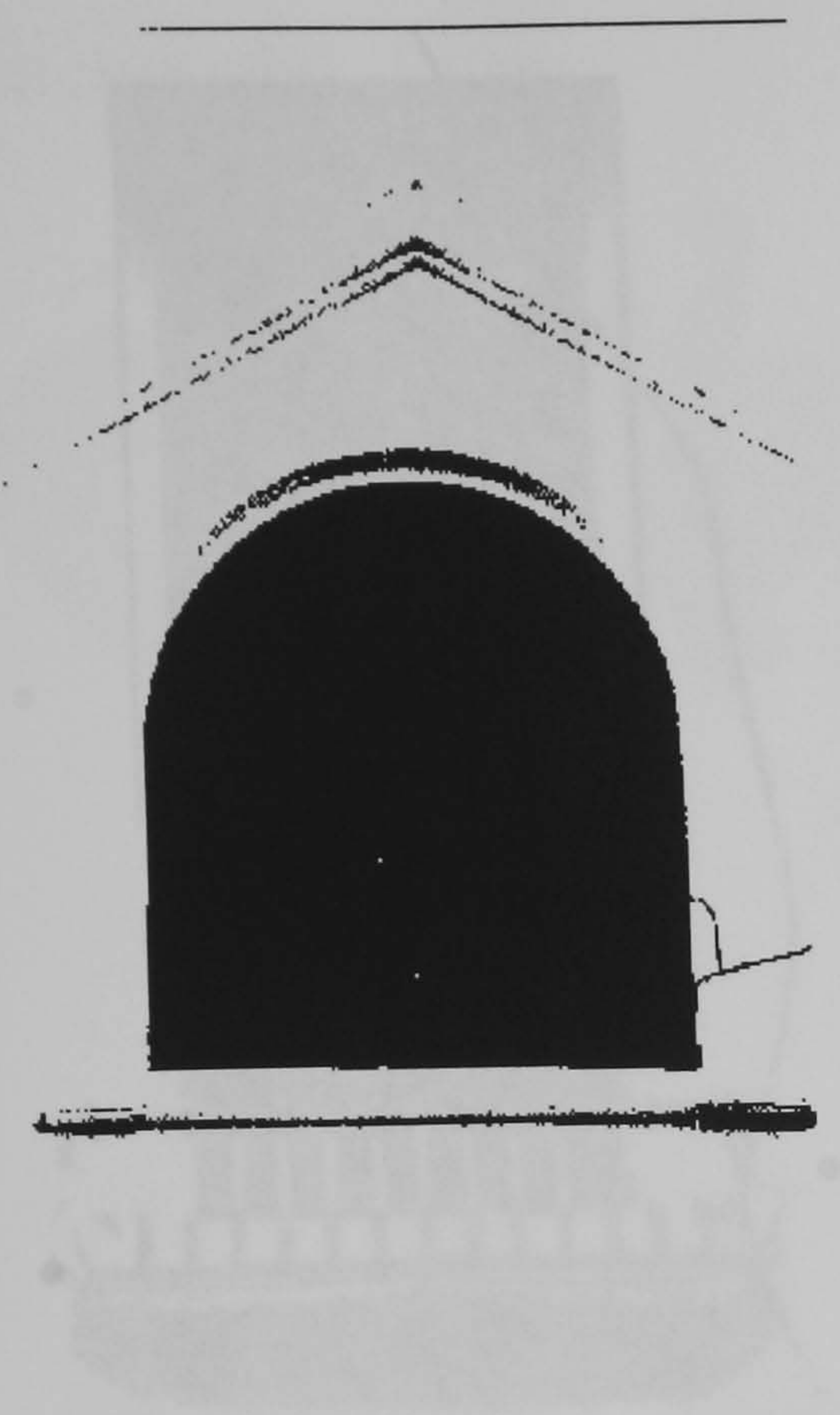


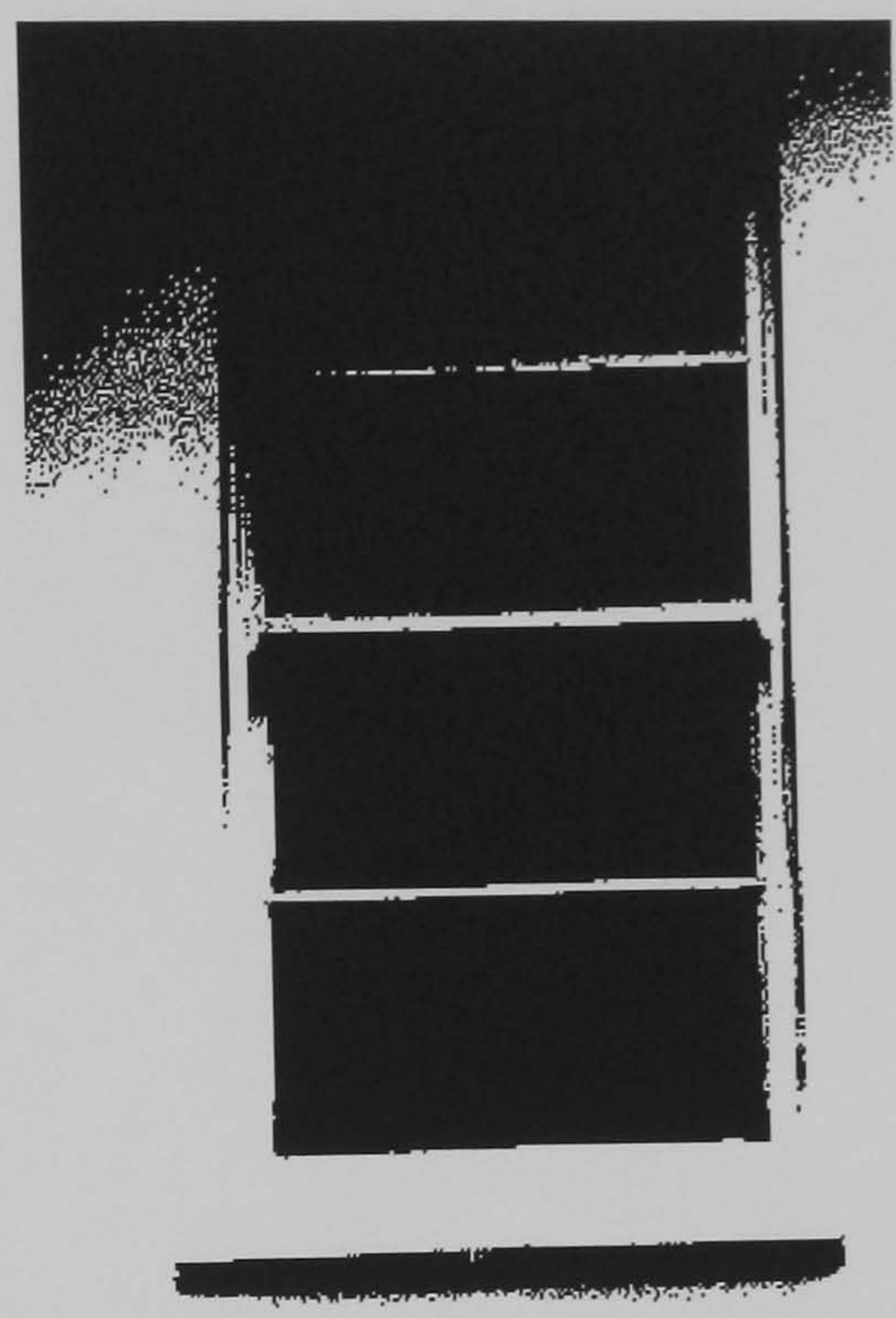


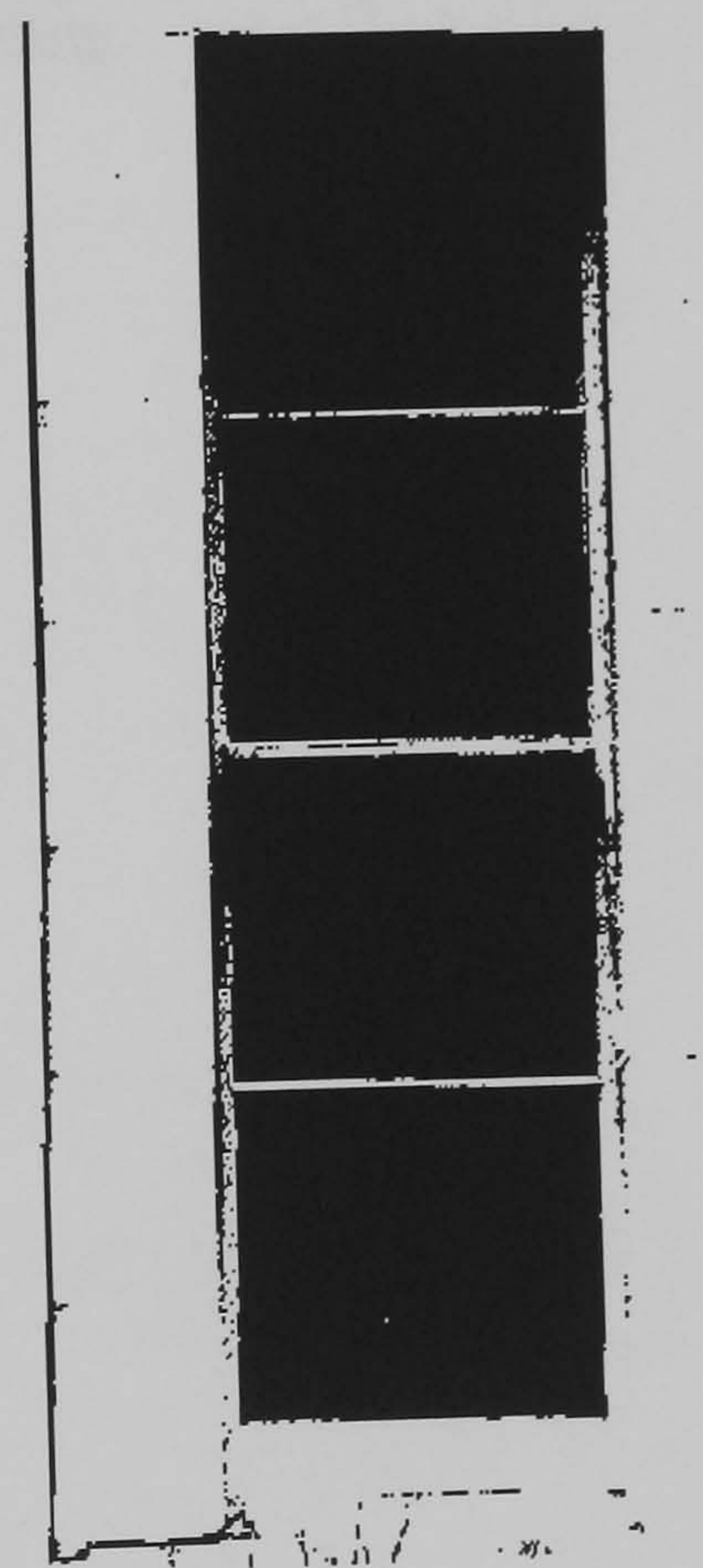
Rather more complex frames

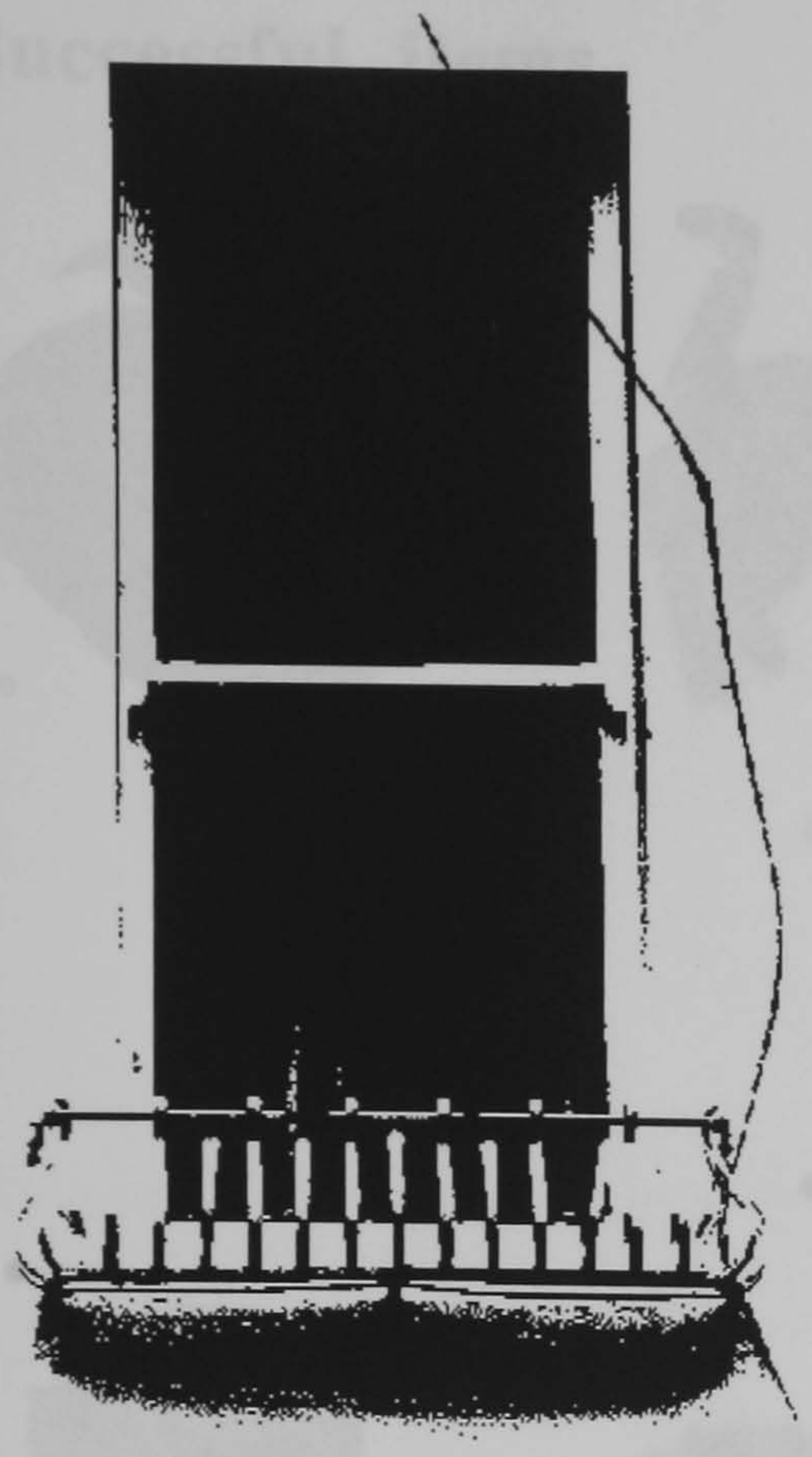










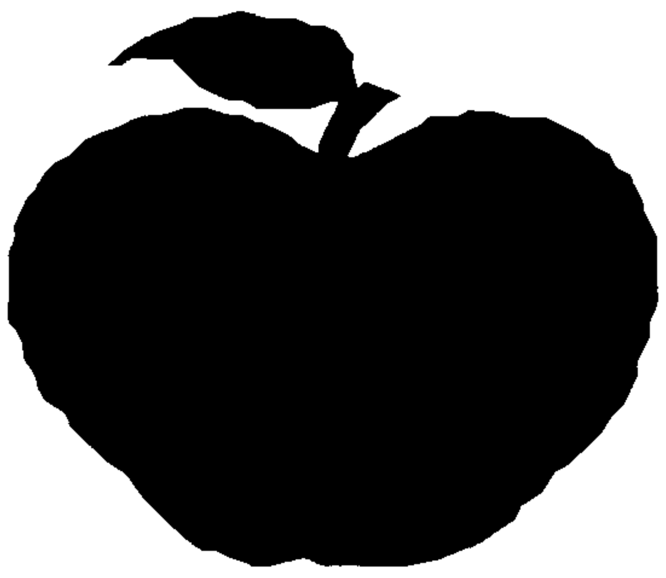


An example of a rejected item: the lead trailing along the side of the window was regarded as being too distinctive and thus open to labelling.

APPENDIX 3

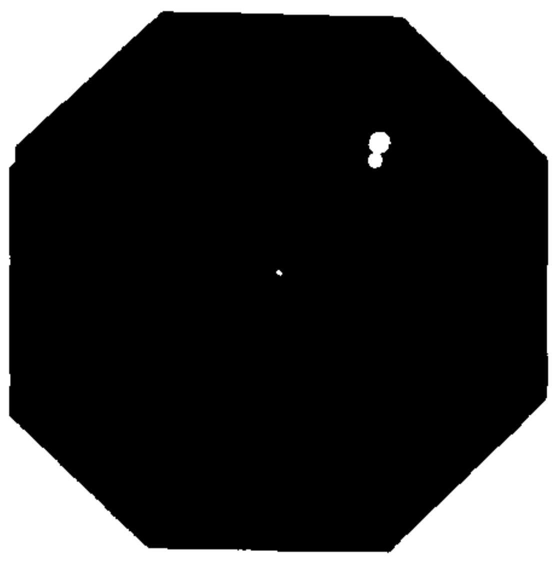
Figure A3. Selected items from the Pictures test

Successful items

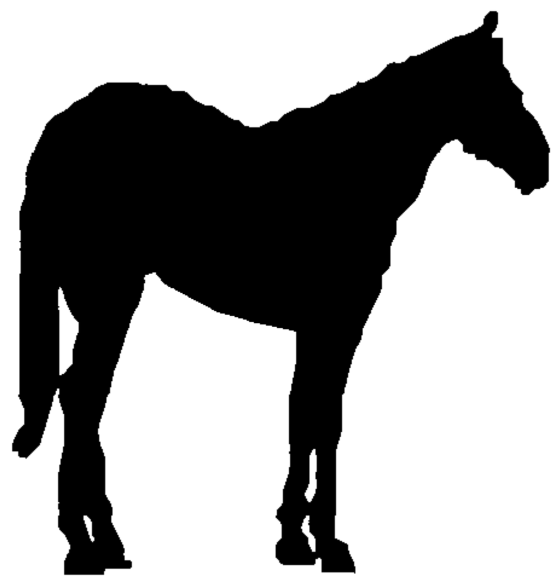


apple

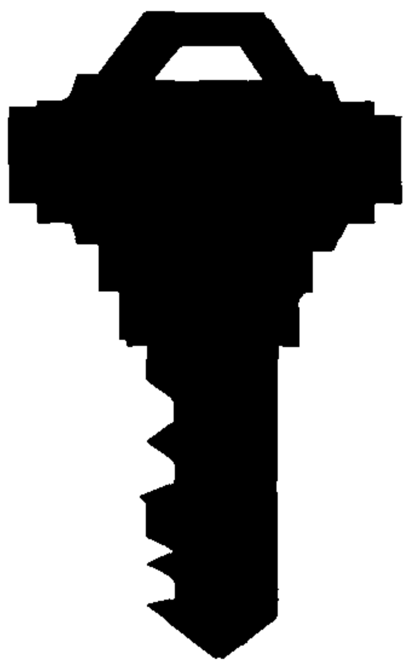
cat



clock



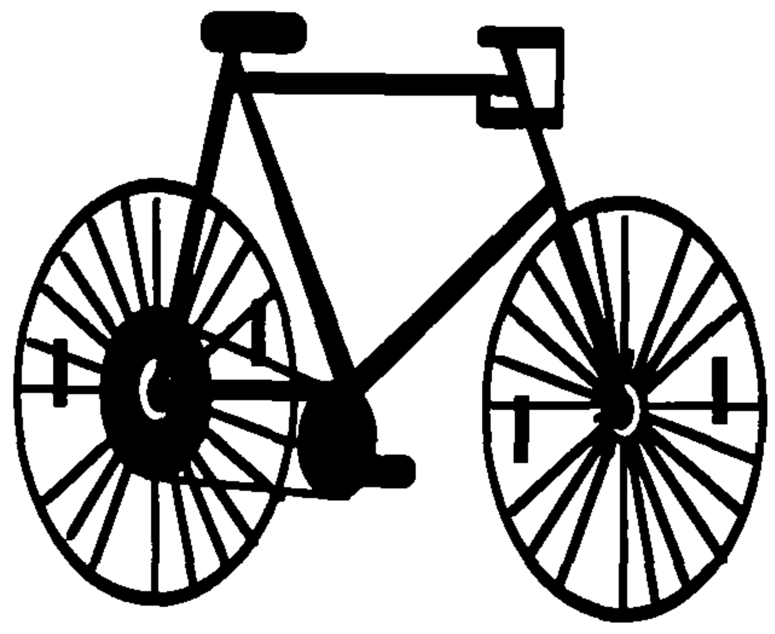
horse



key

tree

scissors

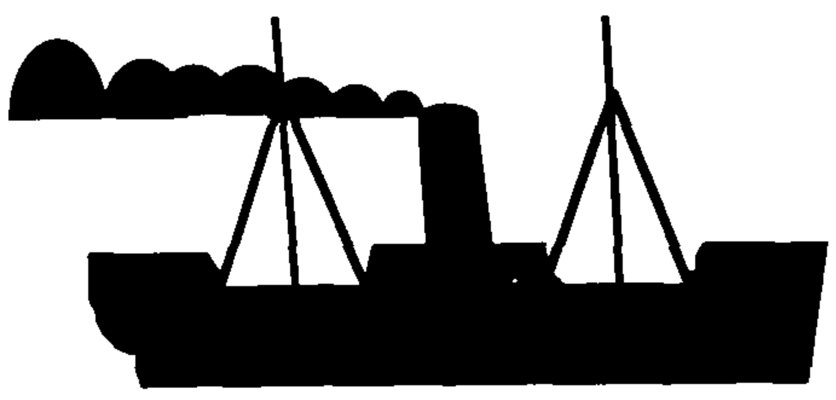


bicycle

Ambiguous items

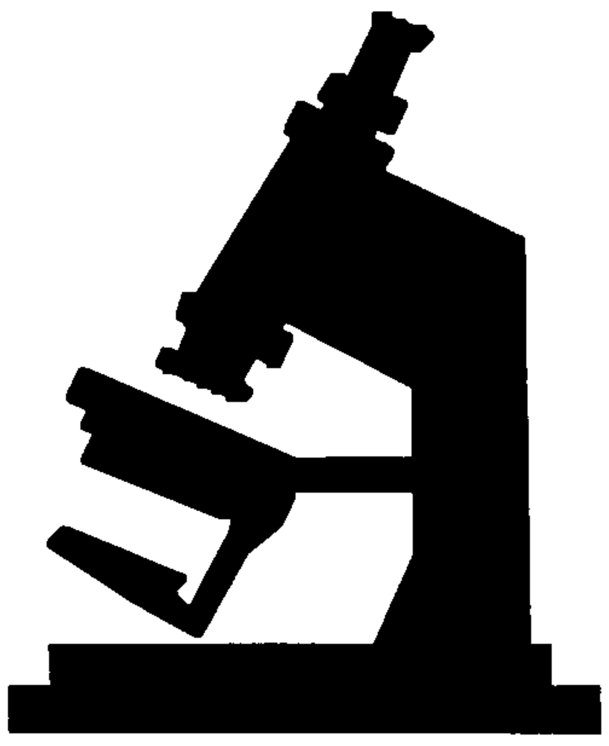
aeroplane/plane

cup/mug

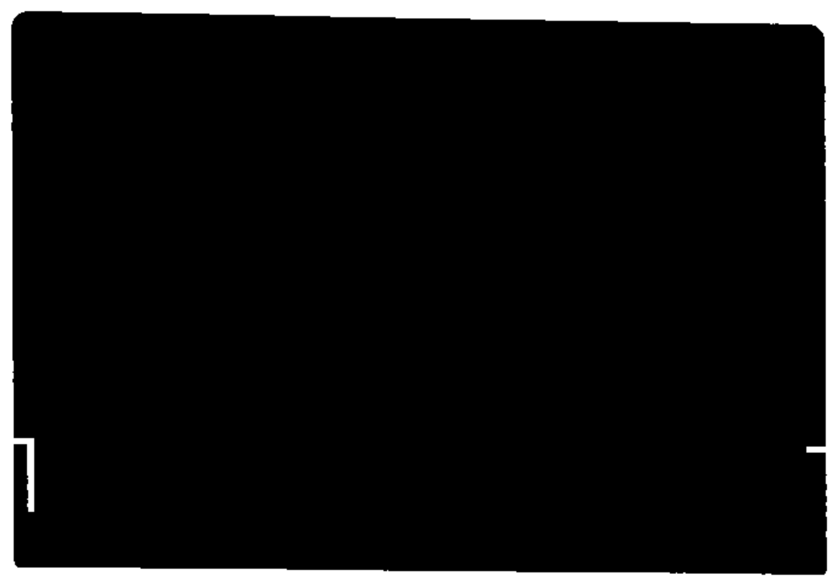


ship/boat

"Technological" items



microscope



video cassette

APPENDIX 4

Words used in Test D of the BAS spelling subscale

morning
eight
friend
know
worse
square
circle
tomatoes
ceiling
armies
conceited
quotation
condition
entirely
occasion
magician
convenient
champagne
gauge
sovereign

APPENDIX FIVE

Mean scores for Chinese and Phoenician spellers in Study 3

	"Phoenicians" (n = 16)	"Chinese" (n = 8)
IQ	86	75
Animals	10.4	9.8
Triangles	9.2	9.6

Mean response times (in seconds) and standard deviations on the three VSM tests in Study 3

	Kirk		Animals		Triangles	
	Mean	SD	Mean	SD	Mean	SD
Controls	3.51	0.7	3.44	0.6	3.52	0.8
Poor spellers	3.44	0.8	3.65	0.7	3.58	0.6